DOCUMENT RESUME

ED 130 263 CS 003 008

AUTHOR	Flynn, Donald L.; And Others
TITLE	An Evaluation of the Cost Effectiveness of Alternative Compensatory Reading Programs, Volume III: Cost-Effectiveness Analysis. Final Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.; RMC Research Corp., Bethesda, Md.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
REPORT NO	RMC-UR-231
PUB DATE	Apr 76
NOTE	166p.; For related documents, see CS 003 002-009; Appendices A and B may reproduce poorly
EDRS PRICE	MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS	*Compensatory Education Programs; Elementary Education; Estimated Costs; *Program Costs; *Program Effectiveness; *Reading Instruction; *Remedial Reading Programs; Test Validity
IDENTIFIERS	Elementary Secondary Education Act Title I; ESEA Title I

ABSTRACT

This report examines the cost effectiveness of reading programs, as part of a large-scale evaluation of compensatory reading programs. Results from a cost analysis of these programs and from an examination of the cost-assessment model used are presented in two other reports. No conclusive relationship was found between the cost of a program and its apparent effectiveness as measured by achievement tests. In addition, with the possible exception of special reading groups, no particular type of class was consistently more effective than another. No unique combination of resources appeared to ensure the success of a program. Limitations and weaknesses in the measure of effectiveness, however, make interpretation of the results difficult. (Author/AA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED130263

RMC Research Corporation
A Resource Management Corporation Subsidiary

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

RMC Report

UR-231

AN EVALUATION OF THE COST EFFECTIVENESS OF ALTERNATIVE COMPENSATORY READING PROGRAMS

Final Report in Four Volumes

Volume III: Cost-Effectiveness Analysis

Donald L. Flynn
Arthur E. Hass
Nabeel A. Al-Salam

April 1976

Prepared for

Educational Testing Service
Princeton, New Jersey

CS 003 008

Home Office: 7910 Woodmont Avenue, Bethesda, Maryland 20014
Other Offices: 2570 West El Camino Real, Mountain View, California 94040
12011 San Vicente Boulevard, Los Angeles, California 90049

CONTENTS

SUMMARY v

1 INTRODUCTION 1

Scope and Study Objectives 1
Background of ETS Study 1
Objectives of Cost Effectiveness 2
Organization of this Report 3

2 UNIT OF ANALYSIS 5

Change in Unit of Analysis 5
Use of Demographic Data 6
Program Clusters 6
Conventional Classrooms versus Special Reading Groups 8
Types of Classes 9
Elimination of Small Classes 11
Class Characteristics 12
Analyses of Variance 17

3 PROGRAM COST 25

Cost-Estimating Model 25
Results of Cost Analysis 26
Analyses of Variance 28
Typal Analysis 34

4 PROGRAM EFFECTIVENESS 41

Measure of Effectiveness 41
Calculation of the Measure of Effectiveness 44
Tests Used for Pre-Test and Post-Test Scores 45
Linear Form of Effectiveness Measure 45
Quadratic Form of Effectiveness 50
Limitations of the Measure of Effectiveness 53
Presentation of Measures of Effectiveness 61

5 MEASUREMENT OF COST EFFECTIVENESS 63

Cost Effectiveness 63
Comparison of Costs Across Clusters 65
Explanation of Observed Changes in Effectiveness in Terms of Program Costs 65
Analytical Approach 67

6 RESULTS OF THE COST-EFFECTIVENESS ANALYSIS 73

Correlation Coefficients 74
Regression Equations 79
Analysis of Variance of Effective Programs 82
Independent Analysis of Post-Test Scores 85
Simple Comparison 96
Separate Examination of Unusually Effective Programs 97

7 CONCLUSIONS AND RECOMMENDATIONS 101

Specific Conclusions 101
General Conclusions 105
Methodological Problems 107
Recommendations 108

APPENDIX A: SCATTER DIAGRAMS OF PRE-TEST VERSUS POST-TEST SCORES . . 111

APPENDIX B: SIGNIFICANT RESULTS OF REGRESSION ANALYSIS FOR THE QUADRATIC MEASURE OF EFFECTIVENESS 129

APPENDIX C: RESULTS OF THE ANALYSIS OF VARIANCE OF THE COSTS OF EFFECTIVE PROGRAMS 143

SUMMARY

Volume III describes the results of a comprehensive study of the cost effectiveness of alternative compensatory reading programs undertaken by RMC Research Corporation. This study was part of a large-scale evaluation of compensatory reading programs conducted by Educational Testing Service (ETS) for the U.S. Office of Education.

The evaluation study involved a comprehensive examination of compensatory reading programs in 256 public elementary schools across the nation during the 1972-73 school year. For this study, compensatory reading programs were defined as reading instruction given to pupils because they were reading below grade level.

Of the total schools involved in the study, approximately 126 had compensatory reading programs funded by ESEA Title I, 80 had compensatory reading programs funded from sources other than ESEA Title I, and 30 had no identified compensatory reading program. The remaining schools (approximately 34, 28 of which were funded by ESEA Title I) were included because they were considered to be exemplary programs.

In each school, all the students in the second, fourth, and sixth grades were given special pre- and post-tests to determine reading achievement and changes in attitude during the school year. Detailed attendance records were kept of pupil participation in reading classes. In addition, the principal and all teachers of students in the affected grades completed questionnaires describing the demographic characteristics of the students and their own approaches to teaching. Finally, the principal completed a questionnaire that identified and enumerated the resources employed in the school.

RMC's responsibility in this study was to first estimate the cost of the individual programs. Then, using measures of program effectiveness developed by ETS, RMC evaluated the cost effectiveness of each program.

The results of RMC's work are presented in four volumes, each of which is described below:

- Volume I: Cost Analysis--this volume develops and uses the methodology for estimating the cost of compensatory reading programs. The resulting cost estimates are presented and analyzed.
- Volume II: Model Sensitivity--this volume examines the methodology developed in Volume I to assess the impact of alternative cost allocation rules and to measure errors.
- Volume III: Cost-Effectiveness Analysis--this volume examines the cost effectiveness of alternative compensatory reading programs using the costs developed in Volume I.
- Volume IV: Cost and Cost-Effectiveness Analysis of Summer Programs--this volume uses the procedures and methodologies developed in the first three volumes to conduct an independent examination of summer programs.

This summary describes the major conclusions and results presented in this volume.

GENERAL CONCLUSIONS

The cost-effectiveness analysis attempted to answer two major questions:

- If functional relationships exist between the cost of program resources and student achievement gains, what is the nature of these relationships?
- Are certain compensatory reading programs more cost effective (i.e., a greater return in achievement gain for each dollar spent) than others?

Unfortunately, the answers that resulted from this analysis were not conclusive. As will be summarized later, the limitations and weaknesses in the measure of effectiveness developed for this study restricted our ability to accurately interpret the results. However, there appeared to be no consistent relationship between the cost of a program and its apparent effectiveness as measured by achievement tests. In addition, with the possible exception of special reading groups, it was shown that no particular type of class was consistently more effective than another. No unique combination of resources appeared to ensure the success of a program.

In fact, when special reading groups were excluded, the treatment received by compensatory students in classes by themselves appeared to be essentially

identical to that received by compensatory students in classes that also contained noncompensatory students.[1] However, the classes with compensatory students only were 15 and 45 percent more expensive, implying that the most cost-effective programs were those that cost the least. Thus, there is little justification for concluding that the more resources a program has, the more effective it is.

This is not to say that costs can be completely ignored, since instruction cannot take place if the resources are below some critical point. However, it is not possible to identify this critical point with the data collected for this study. There did not appear to be any classes that were both ineffective and "resource poor." This finding may be a result of the relationship between the availability of resources and low achievement caused by the impact of ESEA Title I funds. This relationship clouded the issue of a critical resource level so that no firm statement can be made.

The only possible exception to the general rule that program effectiveness does not seem to be influenced by the quantity of resources available is the special reading group. The special reading group is a small group of students--typically 8 to 12--who receive reading instruction from a reading specialist in addition to the reading instruction they receive in the regular or conventional classroom--typically with 20 to 30 students. The gains scored by these groups were generally slightly higher than those achieved by the more conventional compensatory programs. However, the special reading groups were quite expensive, generally two or three times the cost of the conventional classroom.

The question of why there is no consistent relationship between the cost of a program and its apparent effectiveness is not really addressed in this study. However, we can conjecture possible reasons. First and most obvious, the availability of a resource may be less important than how it is used. For example, teacher aides vary from little more than clerks to important extensions of the teacher's ability. In addition, other factors may affect student achievement more significantly than the availability of resources as measured in dollars and cents. Factors such as parent interest, teacher

---

1. This result differs somewhat from ETS findings, where compensatory students in mixed classes performed better than compensatory students alone. However, ETS did not separate special reading groups from the rest of the classes as did this analysis. Thus, the results are not directly comparable.

and staff attitudes, or school environment may be far more important. While it may be true that the availability of resources is indicative of these factors, or actually influences them, there do not appear to be any consistent relationships indicating that greater levels of resources improve a program.

BASIS OF ANALYSIS

The basic unit of analysis used in this study was the class or instructional group. Both an estimate of the cost and a measure of the effectiveness of reading instruction were developed for each class included in the study. The use of the class as the basic unit of analysis made possible by the extensive data-collection effort allowed a more intensive examination of the educational process than is possible at either a program or school level of analysis.

It was possible to identify particular students in the school who were reading below grade level. These students were called compensatory (CR) students. The rest of the students in the school were called noncompensatory (NCR) students. It was decided to use the relative composition of CR and NCR students in a class as a basis for conducting the evaluations performed in this analysis. Because of the differences in student characteristics, it was decided to examine the CR and NCR students separately when both types were in the same class. In addition, it was decided to separate special reading groups from the more conventional classrooms. As defined above, special reading groups are small groups of students--typically 8 to 12--who receive reading instruction from a reading specialist in addition to the reading instruction they receive in the regular or conventional classroom--typically with 20 to 30 students. Using these distinctions, the following seven basic types of classes or instructional groups were defined:

(1) CR students in conventional classrooms alone,

(2) CR students in conventional classrooms together with NCR students,

(3) NCR students in conventional classrooms together with CR students,

(4) NCR students in conventional classrooms alone,

(5) CR students in special reading groups alone,

(6) CR students in special reading groups together with NCR students, and

(7) NCR students in schools with no identified compensatory program (called NCR schools).

A summary of the average enrollment, average pre-test score, and average post-test score for all grades is presented in Summary Table 1. Because different tests were used in each of the grades, separate test results are presented for each grade. A series of analyses of variance was conducted on these characteristics in an attempt to identify whether any significant differences existed between the various types of classes. The results of these analyses are summarized below:

- There were generally no significant differences in enrollment, pre-test scores, and post-test scores between CR students by themselves and CR students in classes with NCR students. This was true for both conventional classrooms and special reading groups.
- Significant differences existed between conventional classes and special reading groups. The enrollment of special reading groups was generally much smaller than it was for conventional classrooms--approximately 10 versus 20. In addition, both the pre- and post-test scores were lower for special reading groups--typically 5 to 10 points lower.
- Classes with NCR students only were generally larger by one or two students than conventional classrooms with a mix of CR and NCR students.
- Classes in NCR schools generally had average pre- and post-test scores that were lower than other conventional classrooms with NCR students.
- Classes with NCR students only generally had lower average scores than NCR students in conventional classrooms with a mix of CR and NCR students.

COST ANALYSIS

Most previous cost studies of education programs have relied on budget reports from the individual school as the primary data source for estimating costs. However, these budget-based estimates have proved inadequate for comparing alternative programs--primarily because the varied accounting procedures used in school districts make it difficult to compare costs. Further, such budgets generally do not highlight the concentrations of resources being examined. Finally, budgetary costs, because they indicate only the current year's expenses, neglect the value of resources inherited from previous years and distort the costs by including one-time capital expenditures.

Summary Table 1

AVERAGE CHARACTERISTICS BY CLASS TYPE[a]

		Compensatory Students				Noncompensatory Students		
		Conventional Classrooms[b]		Special Reading Groups		CR Schools		NCR Schools
		Only	Mixed	Only	Mixed	Only	Mixed	
Average Enrollment[c]		26.5 (5.5)	26.7 (5.1)	8.7 (5.4)	12.5 (6.8)	28.5 (5.0)	26.9 (5.0)	26.8 (5.5)
Grade 2[d]	Average Pre-Test Score	59.6 (17.8)	59.4 (11.8)	54.3 (13.3)	54.9 (10.3)	82.6 (18.6)	88.0 (15.9)	77.5 (15.7)
	Average Post-Test Score	89.6 (16.8)	91.8 (14.3)	84.1 (19.0)	88.2 (16.3)	105.8 (15.0)	111.5 (9.6)	104.0 (11.7)
Grade 4[d]	Average Pre-Test Score	65.8 (17.7)	64.8 (14.6)	54.9 (10.9)	56.9 (8.1)	96.1 (20.4)	99.0 (16.1)	88.5 (16.0)
	Average Post-Test Score	82.1 (17.7)	83.6 (16.4)	73.7 (16.5)	76.8 (13.0)	110.5 (16.7)	112.7 (13.8)	104.1 (15.2)
Grade 6[d]	Average Pre-Test Score	83.3 (21.6)	86.1 (18.9)	66.5 (5.6)	74.3 (17.2)	111.4 (18.7)	117.1 (15.3)	112.5 (12.1)
		93.6 (20.4)	96.1 (18.0)	80.3 (19.2)	85.2 (19.9)	118.4 (16.6)	122.9 (13.9)	118.0 (11.6)

a. Standard deviations are displayed in parentheses below the means.

b. As described in the text, conventional classrooms are considered as the typical classroom setting with one teacher teaching all subjects to 20 to 30 students.

c. Averaged across grades.

d. It should be remembered that different tests were given to the various grades. Thus, the mean values cannot be compared directly.

The method adopted in this study, called the resource approach, was based on a rigorous assessment of the resources (i.e., staff, facilities, equipment, and materials) that were actually used in alternative compensatory reading programs. Specific attention was given to identifying where and how resources were concentrated by collecting all data at the classroom level. A total of 97 resources--including 39 types of staff members--was identified and handled separately in a computerized cost-estimating model.

To make the costs truly comparable, regional variations were eliminated by using average national values as the standardized resource cost (e.g., the salary of a teacher, including fringe benefits), and a proportional amount of the initial capital cost of acquiring the resource (e.g., the purchase price of a tape recorder). All resources used in a program were considered and costed, even resources such as facilities that are generally available at no real cost to the school district. This was done to develop the total cost of the program, which is more comparable between school districts because of the variability in what is available at no cost to school districts.

The procedure used in the cost-estimating model involved estimating the quantity of each resource available to a classroom. Using the standardized costs for each resource, the model calculated the total cost of employing that resource for reading. The total costs were then divided by the number of students in the class to yield an estimate for the average cost per student for each resource. Although the model examined 97 separate resources, it would have been difficult to analyze and interpret the results at this level of detail. Therefore, the costs were aggregated into 20 cost categories based on the function of the resources. An additional five cost categories were formed by aggregating the original 20 cost categories into higher levels. A more detailed discussion of this methodology can be found in Volume I of this report.

A summary of the average costs, in dollars per student, for each type of class is presented in Summary Table 2. This table, which averages all grades, presents both the total cost for all resources and the costs for the major resource components. A series of analyses of variance was conducted to examine the differences in total cost for the various types of classes. The results of these anlayses are summarized below:

Summary Table 2

AVERAGE COST IN DOLLARS PER STUDENT[a, b]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[c]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Class Instructor	68.60 (38.28)	62.72 (43.69)	64.18 (40.65)	74.30 (53.82)	61.07 (36.77)	61.67 (41.04)	63.80 (30.58)
Teacher Aide	16.58 (38.29)	7.74 (20.82)	34.47 (49.97)	25.80 (44.18)	4.44 (10.43)	6.73 (15.25)	8.90 (18.93)
Reading Specialists	30.95 (58.32)	15.76 (25.29)	293.16 (202.75)	223.18 (120.35)	9.63 (16.89)	16.16 (26.76)	10.58 (26.54)
Other Staff	25.50 (24.93)	18.94 (20.43)	43.49 (43.90)	54.77 (75.77)	15.18 (24.26)	18.59 (20.17)	14.99 (13.15)
All Staff	141.63 (90.85)	105.17 (63.69)	435.30 (220.49)	378.05 (186.62)	90.33 (56.87)	103.17 (59.04)	98.26 (47.91)
Facilities	34.11 (18.40)	29.86 (18.70)	161.54 (87.27)	136.62 (58.55)	29.39 (18.23)	29.42 (17.65)	30.98 (13.64)
Equipment	3.45 (6.23)	2.83 (2.73)	9.51 (7.90)	8.97 (7.12)	2.39 (3.01)	2.72 (2.62)	3.78 (2.84)
Materials	11.49 (16.29)	12.23 (5.16)	25.54 (12.28)	27.26 (13.08)	11.05 (8.86)	12.19 (5.64)	12.40 (17.26)
All Resources	196.33 (112.67)	153.18 (80.64)	637.67 (314.45)	555.01 (246.57)	136.68 (75.41)	150.22 (74.74)	148.27 (62.49)

a. Standard deviations are displayed in parentheses below the means.

b. Costs, in dollars per student, are averaged across all grades.

c. For definition of conventional classrooms set Note d, Summary Table 1.

- Generally, there were no differences in cost between grades for the same type of class. However, there were two exceptions to this general rule. The second-grade classes with either NCR students alone, or in NCR schools, were significantly more expensive than the fourth- or sixth-grade classes. The costs of these classes were approximately 24 and 35 percent more, respectively.
- Conventional classrooms with CR students only were generally more expensive than any other type of conventional class, averaging 35 percent more.
- Conventional classrooms with NCR students--whether alone, in mixed CR/NCR classrooms, or in NCR schools--generally cost about the same.
- Special reading groups were significantly more expensive than conventional classrooms. On the average, special reading groups cost approximately 200 percent more than conventional classrooms with CR students only, and 300 percent more than conventional classrooms with a mix of CR and NCR students.
- Special reading groups with CR students were not significantly more expensive than special reading groups with a mix of CR and NCR students.

MEASURE OF EFFECTIVENESS

A meaningful measure of effectiveness for use in a cost-effectiveness study must not only identify which programs are more effective, but must also indicate how much more effective these programs are. The measure of effectiveness chosen for this study was the residual gain score. The residual gain score is the difference in the actual average post-test raw score for a class and the raw score that the students would be expected to achieve considering their initial skill levels and experience. In algebraic form, the measure of effectiveness is calculated as follows:

$$\text{effectiveness} = \text{actual post-test} - \text{expected post-test}.$$

The theoretical advantage of a residual gain score used as a measure of effectiveness is that it allows meaningful comparisons to be made at different levels of initial skill and experience, while compensating for the nonlinear relationship between true achievement and the test score. This nonlinearity precludes the use of simple post-test minus pre-test scores as a measure of effectiveness.

The procedure followed in calculating the expected post-test score involved estimating a relationship between pre-test and post-test raw scores to minimize the variance (i.e., the square of the difference between actual post-test and expected post-test for any class). The functional form used in this procedure, called regression analysis, took the quadratic form, or:

$$\text{expected post-test} = a_1 + a_2\,(\text{pre-test}) + a_3\,(\text{pre-test})^2,$$

where a_1, a_2, and a_3 are the equation coefficients. A regression analysis was conducted for each class type in each grade. Unfortunately, this procedure forced the average effectiveness for each type of class to zero. Thus, it was not possible to directly compare the effectiveness of the various class types.

Limitations of the Measure of Effectiveness

In theory, the residual gain score should have proved to be a meaningful measure of effectiveness. However, the measure of effectiveness had two inherent weaknesses that significantly reduced its usefulness in the cost-effectiveness analysis.

The first was the use of test scores for all students who were ever enrolled in the class, even if both pre-test and post-test scores were not available. The number of situations where only one score was available did not represent a large portion of the total observations. Nevertheless, such situations introduced effects that were neither the result of the educational approach nor a function of the initial class achievement level. Although these effects are not readily apparent in the data, their presence tends to reduce the overall confidence in the results.

The second weakness of the measure of effectiveness lies in its general inability to provide a basis for meaningful comparisons between classes with different pre-test scores. In particular, the interval scale used to interpret the measure of effectiveness varied with the pre-test score. It was shown that the sensitivity of the measure decreased by a factor of five between pre-test scores of 50 and 100 for the second-grade test. This weakness created analytical problems that made it nearly impossible to detect relationships between cost and effectiveness.

COST-EFFECTIVENESS ANALYSIS

Despite the limitations in the measure of effectiveness, a concerted effort was made to examine all meaningful relationships between cost and effectiveness. Alternative definitions of what constitutes effective programs were examined in an effort to explore all possible avenues. A total of six such approaches was used. Each of them is discussed below, along with its corresponding results.

Correlation Coefficients

The correlation coeffficient, called R, is a measure of how closely the relationship between two variables approaches a linear relationship. The coefficient, which can range from -1 to +1, is useful in a cost-effectiveness analysis as a means of determining the correlation between cost and program effectiveness. If the R calculated between cost and effectiveness is greater than zero, higher costs are generally associated with higher levels of effectiveness. Conversely, if the R is less than zero, higher costs are generally associated with lower levels of effectiveness. The closer R is to either -1 or +1, the stronger the relationship between cost and effectiveness. When R is zero, there is no relationship between cost and effectiveness.

The results of the analysis of correlation between cost and the measure of effectiveness calculated using the quadratic form did not suggest any clear relationship between cost and effectiveness that was consistent over a variety of situations. Very few coefficients were greater than an absolute value of 0.25. At this level, less than 7 percent of the variation between effectiveness and cost could be explained by a relationship between the two. Of additional concern was the frequency of negative correlation coefficients. Approximately half the significant relationships were negative. A negative coefficient is difficult to explain because it implies that higher levels of cost are associated with less effective programs. This, however, might be caused by a third factor that may be influencing both the cost and the effectiveness. A prime example of this may be socio-economic status. It is possible that students with lower socio-economic status are more likely to score lower on the tests. However, because of their lower socio-economic status, they are eligible for additional federal funds under ESEA Title I. Thus, low effectiveness and high cost may be associated because of the presence of ESEA Title I funds.

Regression Equations

In an attempt to identify an explicit relationship between the various categories of cost and the measure of effectiveness calculated using the quadratic form, multiple linear regression techniques were used with the following functional form:

$$PE = C + \Sigma a_i X_i ,$$

where

PE = program effectiveness,

X_i = program cost variables for each particular resource type, and

C, a_i = regression coefficients.

The regression coefficients represent the marginal (or incremental) increase in program effectiveness that can result if one additional dollar is spent in providing that resource. However, this interpretation may not be completely appropriate for this analysis.

The results for this analysis indicate that there was very little consistency in significant variables across the various class types. Seldom was any particular resource cost category found to have a statistically significant relationship with effectiveness in more than one of the 11 different equations developed. However, in three cases--

- cost of teacher aides,
- cost of basal readers, and
- cost of district centers--

the costs were significant in more than one class type. The results were somewhat consistent with respect to the analysis of correlation--negative for teacher aides, positive for basal readers, and usually negative for district centers.

In general, however, the regression equations explained very small amounts of the total variation in the measure of effectiveness, averaging only 9 percent. Thus, the result tends to support the basic conclusion that there is no relationship between cost and effectiveness.

Analysis of Variance of Effective Programs

The proper use of multiple linear regression techniques requires a great deal of statistical integrity that, as we have already pointed out, may be lacking in the data. In particular, the problems of the nonlinearity and heteroscedasticity in the estimation of effectiveness may make the results of the linear regression analysis meaningless.

In an attempt to overcome these problems, an alternative approach to evaluating the cost effectiveness of various resources was tried that did not depend as heavily on the measure of effectiveness. This approach involved stratifying the individual classrooms into separate cells by pre-test and post-test scores. The classrooms were stratified by pre-test score to allow an examination of the impact of the pre-test score on cost effectiveness. The stratification by post-test score was accomplished by subjectively dividing the post-test scores for each pre-test stratum into the post-test scores that could be considered better than average, those that were about average, and those that were below average. Once the strata were defined, analysis of variance techniques were performed to determine if significant differences existed in the costs associated with each of the cells.

Many contradictions can be seen between the results from this analysis and those from the regression equations. These contradictions cannot be resolved because two separate approaches were used with different objectives. Despite the apparent conflicts, however, both analyses support the argument that significant relationships between cost and program effectiveness do not exist. Both analyses show very low values for the R^2 statistic, indicating that other factors influence the effectiveness of a class far more than the quantity of resources available. Therefore, it is entirely possible that spurious results are responsible for the presence of false relationships between particular cost categories and effectiveness. This, in fact, could explain the apparent contradiction between the two analyses.

Independent Analysis of Post-Test Scores

As another approach, the relationships between cost and program effectiveness were examined by dividing each type of class into groups representing high, medium, and low total costs. The expected post-test scores were calculated by regression analysis for each category using the quadratic

form discussed earlier. The expected post-test scores for all three groups were plotted to determine which group had the highest post-test score at each level of pre-test.

In general, no significant differences were found between the post-test scores for classes at the various cost levels. This supports the previous findings of no significant relationship between cost and program success. However, there is an exception to this generalization. Generally, higher post-test scores were found in special reading groups when compared with conventional classrooms.

Simple Comparison

As a final examination, we made a systematic review of the various characteristics, cost, and effectiveness data. The major conclusion reached in this analysis was that CR students--both alone and in mixed classes--appeared to achieve the same educational gains starting from the same pre-test levels. However, the cost associated with conventional classrooms with CR students only was between 15 and 45 percent higher than conventional classrooms with a mix of CR and NCR students. Again, this supports the general conclusion that there is no relationship between cost and program success.

While this study indicates where the additional money is being spent (in this case, the additional costs are associated with increased use of teacher aides and reading specialists), it is not possible to explain why it was spent. Obviously, other factors affect cost and effectiveness that have not been fully explored in this study. Some of these factors may involve value judgments made by local school administrators based on variables not considered in this study. While programs funded by ESEA Title I were demonstrated to cost more than other programs (see Volume I), the availability of ESEA Title I funds does not appear to be one of the factors that influence the higher costs associated with conventional classrooms having CR students only.

A Separate Examination of Unusually Effective Programs

In-depth observations were conducted by ETS at 29 schools in the sample that displayed various levels of effectiveness to determine the specific characteristics of their programs. Neither the trained observors nor the schools were informed of the school's rank in terms of effectiveness.

Five schools were identified as being unusually effective. This selection was based on the observational data and the test results. An analysis of the total costs of the classes in these five schools indicated no significant differences between them and the other schools included in the analysis. The fact that the costs of effective programs are not significantly different from the costs of average programs lends support to the general conclusion that effectiveness is not functionally related to the resources available to a compensatory reading program (and thus its costs).

RECOMMENDATIONS

Because of the lack of conclusive results, no specific recommendations could be made concerning educational policy. It is true that no relationships were uncovered between cost and program effectiveness. However, this is not conclusive proof of no relationship. Our inability to find such a relationship may be the direct result of the limitations in the data previously described. In an attempt to overcome these limitations in future studies, several recommendations have been made in this volume, which are summarized below:

(1) Before any future studies are undertaken, the problems associated with a meaningful measure of effectiveness should be examined in detail. In particular, the variation in test sensitivity should be examined. Solutions that may include the use of correction factors or the use of scaled scores should be developed. If these or other solutions are not feasible, consideration should be given in the analysis plan to restrict the examination to a series of relatively homogeneous treatment groups with respect to pre-test scores.

(2) Future analyses should ensure that a series of definitive educational approaches be used as the basis for the analysis.

(3) Future analyses should examine in greater detail the impact of policy decisions on the allocation of resources. At the very least, future analyses should include an examination of the relationship between the socio-economic levels of the students and the level of resource funding.

1

INTRODUCTION

SCOPE AND STUDY OBJECTIVES

This report is the third volume of a four-volume study on the cost-effectiveness of alternative compensatory reading programs in selected elementary schools throughout the country. The entire study was designed to complement and support a large-scale evaluation of such programs by the Educational Testing Service (ETS). Volumes I and II of this study developed and evaluated a cost-estimating model to determine the costs of these programs, and actually estimated those costs. This report integrates the costs with the program outcomes to evaluate the cost effectiveness of the compensatory reading programs.

BACKGROUND OF ETS STUDY

In July 1971, ETS began a survey of public elementary schools and selected features of their reading programs. The study was conducted in three phases: The first phase involved a descriptive and analytical study of reading programs in grades 2, 4, and 6 in 731 schools throughout the United States. Phase II included a comprehensive evaluation of the programs for a sample of 256 schools during the 1972-73 school year. The goal of Phase II was to determine how students benefitted from various reading methods and categories of instruction. In addition, the cost effectiveness of the various methods was examined. Phase III involved a similar evaluation for special summer compensatory reading programs in 27 schools during the summer of 1973.

ETS used a number of instruments for recording the necessary data and descriptions of the schools and reading programs. The instruments included standardized tests to determine reading achievement and student attitudes

toward reading, class attendance records, and principal and teacher questionnaires for collecting program and demographic characteristics.

While the principal and teacher questionnaires determined the use of various types of resources (staff, equipment, facilities, and materials) in the reading programs, no attempt was made to quantify the amount of resources available to the teachers. For this reason, in October of 1972, ETS asked RMC to develop a supplemental resource survey and to conduct the cost-analysis and cost-effectiveness evaluations for the Phase II study. Volumes I, II, and III of this report present the results of RMC's contribution to Phase II. In August of 1973, RMC began the cost analysis and cost-effectiveness analysis of the summer programs under Phase III. Volume IV of this report presents the results of that examination.

OBJECTIVES OF COST EFFECTIVENESS

In the United States today, it is not sufficient to examine only the effectiveness of alternative compensatory programs. In light of constant budgetary constraints and the resultant difficulty in obtaining additional funding, it is also necessary to examine the cost of these programs and the relationships between their cost and effectiveness. Thus, the primary objective of this study is to examine the relationships that exist between cost and effectiveness.

Where alternative compensatory reading programs can be identified, their cost effectiveness will be evaluated. More importantly, the impact of resources, and thus cost, on effectiveness will be evaluated. Does the effectiveness of a particular program increase more with an increase in the availability of tape recorders or an equal cost increase in the availability of teacher aides? What combinations of resources interact to produce more effective programs? These kinds of questions are examined in this study.

Historically, most studies of educational cost effectiveness have not been able to identify significant relationships between cost and effectiveness. Most of the blame for this lack of positive relationships has been placed on the lack of comparable measures of effectiveness and cost. For this reason, this study attempted to develop meaningful measures of both effectiveness and cost that could be compared across various classrooms, schools, and districts. The measure of effectiveness was based on the results of identical reading achievement tests administered in all classrooms

at approximately the same time in the beginning and again at the end of the school year. A residual gain score was calculated using raw scores as a measure of program effectiveness. This procedure was used to separate the treatment effects from the initial achievement levels. Because of the variation in budgetary procedures between districts, the costs of instruction were not estimated using published budget costs. Instead, the costs were estimated directly from the resources that were found to be available in the classroom.

ORGANIZATION OF THIS REPORT

The unit of analysis used in this study of cost effectiveness is the class. This level of analysis allows examination of the actual reading instruction provided to students in each class. If the analysis had been conducted at a higher level, say a program or a school, the class-to-class variation in approach might well have masked the specific effects of the reading program under consideration. Chapter 2 describes the basic reading programs and class types that served as the framework within which the analysis was conducted. Finally, the characteristics of the various class types are compared and contrasted.

As pointed out earlier, many past studies have been criticized because of the poor quality of the data used as inputs. In general, the low quality of data resulted from the lack of cost information and measures of effectiveness that could be meaningfully compared across schools and classes. However, one of the basic strengths of this study is the effort that was made to collect an extensive data base from which truly meaningful inputs could be developed. As a result of its importance, Chapters 3 and 4 present an in-depth discussion of the data base. Although the cost data were presented and discussed in Volume I of this report, Chapter 3 presents additional cost analyses for the various class types. Chapter 4 examines the strengths and weaknesses of the measures of effectiveness that were developed.

Chapter 5 addresses the meaning of a cost-effectiveness analysis and how its results can be useful to an educator. In addition, it discusses the procedures that were used to conduct the analysis, which involved examination of correlation coefficients and regression analysis. Because of our concern about the weakness of the measure of effectiveness, additional analyses are examined that are more independent of the measure of effectiveness. Essentially,

these analyses examine more homogeneous groups of classes. A total of five independent analyses are discussed in Chapter 5.

Finally, the results of each of the five analyses are presented in Chapter 6. The conclusions and recommendations drawn from an examination of these results are presented in Chapter 7.

2

UNIT OF ANALYSIS

The primary unit of analysis used in this study of cost effectiveness is the class or instructional group. Both an estimate of the cost and a measure of the effectiveness of reading instruction have been developed for each class included in the study. The use of the class as the unit of analysis, made possible by the extensive data collection effort, allows a more intensive examination of the educational process than is possible with either a program or school level of analysis.

Using the class as the unit of analysis represents a slight modification in approach from that originally presented in Volume I of this report. This chapter discusses that modification. In addition, this chapter describes the various types of classes and their program characteristics.

CHANGE IN UNIT OF ANALYSIS

The cost analysis conducted in Volume I of this report also used the class as the basic unit of analysis when the cost estimates were developed. However, in the analysis of these estimates (described in Chapter 4 of Volume I), the class estimates were aggregated into grade and school averages. In addition, the definition of what constituted a compensatory (CR) or a noncompensatory (NCR)[1] class was significantly modified. The definitions used to separate the CR and NCR classes in the cost analysis were derived from a distinction made by the principal about whether or not the class provided any reading instruction to students because they were reading below grade level.

---

1. The NCR classes were called developmental classes in the cost analysis presented in Volume I of this report.

Additional data were also collected about each student in the study regarding whether or not he was receiving reading instruction because he was reading below grade level. It was decided that it would be more appropriate in subsequent analyses to use the student-based definition of the CR/NCR division rather than the general class distinction used before. The implication of this new definition is discussed in a later section.

The changes in the unit of analysis and definition described above do not affect the original cost estimates developed in Volume I. They were developed at the class level without regard to the definition of CR classes. However, the analysis of the cost estimates in Volume I is no longer appropriate because of the change in the definition of CR classes. For this reason, Chapter 3 of this volume conducts a re-examination of the cost estimates using the new definition.

USE OF DEMOGRAPHIC DATA

A great deal of program and demographic data were collected by ETS regarding each class involved in the analysis. However, only limited use was made of them in the analysis of program effectiveness because ETS determined that such descriptors were highly correlated with pre-test scores. As a general rule, low pre-test scores are more likely to occur in conjunction with low socio-economic status. The use of highly correlated demographic characteristics when estimating program effectiveness would create the potentially serious problem of spurious results caused by the intercorrelations.

PROGRAM CLUSTERS

In an effort to define a limited number of reading programs, ETS used factor analysis techniques in its Phase I report to identify a series of program clusters. The clusters were based on five reading program indices: (1) emphasis on basic reading activities, (2) use of audio-visual equipment and materials, (3) emphasis on supplementary reading activities, (4) instructional flexibility, and (5) whether or not the compensatory reading programs were offered during time released from other subjects. A complete discussion of the factor analysis is included in the ETS report.[1] The

1. Rubin, Trismen, Wilder, and Yates, A Descriptive and Analytical Study of Compensatory Reading Programs, Educational Testing Service, PR-73-28 (August 1973).

original analysis identified ten clusters for compensatory reading programs using the five reading indices, one "conglomerate" cluster for nonclassifiable compensatory programs, and one cluster for all schools with no identified compensatory programs. These 12 clusters were used in the cost analysis described in Volume I.

After additional analysis, ETS redefined the program clusters. The first 11 clusters remained the same. Clusters 1 through 10 represented the ten identifiable clusters based on the five reading program indices. Cluster 11 was for nonclassifiable schools. A new cluster 12 was formed of classes in schools that were known to have CR students, but for which no compensatory classrooms were identified. Cluster 13 was for all classes in schools with no identified compensatory programs. A new cluster, cluster 14, was used for schools that had apparently unresolved contradictions in the CR/NCR status of the students.

Based on the average school characteristics determined from an examination of the Program and Class Characteristics Questionnare for that school, each class in the study was assigned to one of the 14 clusters. The number of classes in each program cluster by grade is presented in Table 1.

Table 1

NUMBER OF CLASSES IN EACH CLUSTER BY GRADE

Cluster	Grade 2	Grade 4	Grade 6
1	71	71	50
2	103	97	72
3	50	46	27
4	87	92	131
5	37	27	18
6	33	49	71
7	3	6	5
8	41	36	34
9	23	15	9
10	46	48	41
11	84	91	56
12	25	28	13
13	56	57	41
14	37	31	45

Because of the unresolved contradiction in cluster 14, it was eliminated from all the analyses. This represents a reduction in the data base of approximately 6 percent for which both cost and effectiveness data were available.

CONVENTIONAL CLASSROOMS VERSUS SPECIAL READING GROUPS

Appendix B of Volume I shows that it is possible to identify particular groups that are composed of students receiving instruction from two teachers. Generally, all the students in such groups appear on the enrollment records for two classes. In addition, the total number of students enrolled in these groups was usually much lower than the average class--typically 8 to 12 as opposed to 20 to 30 in the average classroom. As a result, these groups seem to represent special reading groups outside the conventional reading program.

It should be pointed out here that some inconsistencies are associated with the identification of these special reading groups. If the instructor of such a special reading group did not keep enrollment and attendance records using the standardized forms supplied by ETS, no identification was possible. In fact, the majority of the reading specialists identified as resources in the schools did not complete enrollment records, and thus are not associated with a particular special reading group. In such situations, the reading specialist is treated as a general resource available to all regular classrooms in the school. A more detailed examination of this problem, and how it was dealt with in the cost model, can be found in Volume I of this report, beginning on page 19.

Except for their small enrollments, the true nature of the identified special groups is not known. However, it seems reasonable to assume that the majority of these groups represent some type of pull-out program. Students in such groups receive basic reading instruction from a teacher in conventional classrooms and supplemental reading instruction from another teacher. Whether this supplemental instruction is provided through individual tutoring or in small groups is not known because of the manner in which attendance was taken. Regardless of its exact nature, the true cost of providing reading instruction to these students is the sum of the costs per student of both classes in which the students are enrolled. Although the sum was not calculated in the cost analysis presented in Volume I, this correction was made in the current analysis to accurately estimate the total cost of reading instruction provided to these students.

The reading instruction provided to students in special reading groups appears significantly more intense than that provided to most students. Not only do students receive instruction from two teachers, but the supplementary reading instruction is generally provided in small groups. As will be demonstrated shortly, the cost per student for special reading groups is approximately two to three times higher than the cost of conventional classrooms. Therefore, it was decided to separate the compensatory classes into conventional classrooms and special reading groups. This would allow us to examine the impact on achievement of the two different treatments.

TYPES OF CLASSES

As discussed above, the distinction between CR and NCR in this report is at the individual student level. Thus, the following types of classes can be defined, depending on the types of students found in each:

- classes with CR students only,
- classes with NCR students only in schools with CR students,
- classes with a mix of CR and NCR students, and
- classes in schools with no CR programs (called NCR schools).

The number of classes by grade for each of these types of classes is presented in Table 2.

Because of the differences in characteristics and treatment, it was decided to further differentiate these four types of classes by separating special reading groups from conventional classrooms. In addition, it was decided to treat CR and NCR students in mixed classes as separate groups. In several instances, special reading groups were identified that had both CR and NCR students. In such instances, the number of NCR students almost always equalled one, and may have resulted from minor errors on the part of the schools in the definition of CR students. Because of their small number, the NCR students in mixed special reading groups were not examined. However, the distinction between CR students in special reading groups alone and CR students in special reading groups with NCR students was maintained.

This brings the total number of instructional groups or class types examined in the analysis to seven. These class types are listed below:

1. CR students in conventional classrooms alone,[1]

---

1. The typical classroom setting with one teacher teaching all subjects to 20 to 30 students.

Table 2

NUMBER OF CLASSES BY GRADE AND CLASS TYPE[a]

	Compensatory Reading Classes	Noncompensatory Reading Classes in a School that Has Compensatory Reading	Mixed Classes	Classes in Noncompensatory Schools
Grade 2	179	101	323	56
Grade 4	177	94	353	57
Grade 6	128	134	265	41

a. Cluster 14 classes have been eliminated.

2. CR students in conventional classrooms together with NCR students,
3. NCR students in conventional classrooms together with CR students,
4. NCR students in conventional classrooms alone,
5. CR students in special reading groups alone,
6. CR students in special reading groups together with NCR students, and
7. NCR students in NCR schools.

ELIMINATION OF SMALL CLASSES

The measure of effectiveness described in Chapter 4 uses both the average pre-test and average post-test raw score for each class in the study. To ensure statistical integrity in later analyses, these two average test scores are based on the results of at least four students in conventional classrooms. If fewer than four students took either the pre-test or the post-test, that conventional classroom was eliminated from the analysis. Since the typical conventional classroom had between 20 and 30 students, it seemed reasonable to eliminate classes with three or fewer students taking the test because a measure of effectiveness based on so few students is subject to increasing sampling error as class size decreases. As a result of the smaller number of students found in the special reading groups, typically six to eight, a special reading group was accepted if only two students took the tests.

The number of classes and student groups remaining after the small classes were eliminated is presented in Table 3. This table shows the size of the data base that was used in all subsequent analyses for each of seven types of classes by grade.

By comparing Tables 2 and 3, it is possible to determine the number of classes that were eliminated because of their small size. For example, the number of conventional classrooms and special reading groups with CR students is only 165. This means that 14 small classes, or 8 percent, have been eliminated. For all classes except those with a mix of CR and NCR students, the number eliminated averaged 6 percent.

Since the mixed classes were treated as two separate student groups (CR mixed and NCR mixed), each group had to have a minimum of four students taking the test (or two in the case of the special reading groups) to be included in the analysis. This rule eliminated approximately 20 percent of the mixed classes--a much larger number than occurred for the other types of classes.

Table 3

NUMBER OF CLASSES BY TYPE OF CLASS

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[a]		Special Reading Groups[b]		CR Schools		NCR Schools
	Only[c]	Mixed[d]	Only[c]	Mixed[d]	Only[c]	Mixed[d]	
Grade 2	133	226	32	27	89	264	55
Grade 4	125	222	39	25	88	291	56
Grade 6	98	187	22	13	127	222	41

a. Classes with three or less students taking pre-test or post-test have been eliminated. Conventional classrooms are considered as the typical class setting with one teacher teaching all subjects to 20 to 30 students.

b. Classes with only one student taking pre-test or post-test have been eliminated.

c. This tabular format will be used in all subsequent tables to indicate classes or groups with CR or NCR students only.

d. This tabular format will be used in all subsequent tables to indicate classes with a mix of CR and NCR students.

However, it was possible to include part of a mixed class, while the second part was eliminated. For example, if a mixed class had test results on 15 NCR students and 2 CR students, only the NCR group was included in the analysis because the CR group had fewer than four observations.

The possibility of eliminating either the CR or the NCR group in a mixed class, while keeping the other group, removed the strict congruence between the two data sets. Thus, a 100 percent overlap for data dealing with different types of students in the same class is not present. Nevertheless, 90 percent of the CR groups have a corresponding NCR group in the same class in the data base. The equivalent number for the NCR groups is 74 percent. It will be shown in later sections that, while some apparent differences exist in the enrollment and cost between the two sets of classes, there are no significant statistical differences. Thus, it seems reasonable to treat the CR and NCR students in mixed classes as if they were in different parts of the same class.

CLASS CHARACTERISTICS

Tables 4, 5, and 6 present a summary of the average characteristics for each type of class in each grade. These tables show the average enrollment, the average number of students who took at least one test, the average pre- and post-test scores, and the percentage of classes in schools receiving full or

Table 4

AVERAGE CHARACTERISTICS BY CLASS TYPE FOR GRADE 2[a]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[c]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Average Enrollment	25.3 (5.1)	25.8 (4.7)	8.5 (3.8)	13.2 (5.8)	26.8 (4.8)	25.9 (4.8)	25.1 (4.0)
Average Number of Students Taking Tests by Type of Student	21.8 (10.5)	12.0 (7.5)	9.2 (3.9)	13.3 (5.9)	26.6 (7.8)	17.2 (7.6)	23.2 (7.1)
Average Pre-Test Score	59.6 (17.8)	59.4 (11.8)	54.3 (13.3)	54.9 (10.3)	82.6 (18.6)	88.0 (15.9)	77.5 (15.7)
Average Post-Test Score	89.6 (16.8)	91.8 (14.3)	84.1 (19.0)	88.2 (16.3)	105.8 (15.0)	111.5 (9.6)	104.0 (11.7)
Percentage of Classes in Schools Receiving Full Title I Funding for CR Program[b]	28	34	69	74	53	31	0
Percentage of Classes in Schools Receiving Full or Partial Title I Funding for CR Program[b]	74	58	94	84	71	60	0

a. Standard deviations are displayed in parentheses below the means.

b. Approximately 38 percent of the schools did not indicate the source of funds for their compensatory programs. The numbers on this line are based on the responses of those schools that indicated a source of funds.

c. As described in the text and used in all analyses, conventional classrooms are considered as the typical classroom setting with one teacher teaching all subjects to 20 to 30 students.

Table 5

AVERAGE CHARACTERISTICS BY CLASS TYPE FOR GRADE 4[a]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[c]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Average Enrollment	27.1 (5.5)	27.0 (5.0)	9.8 (6.8)	11.2 (4.7)	29.2 (5.0)	27.4 (4.7)	27.2 (5.4)
Average Number of Students Taking Tests by Type of Student	23.2 (8.4)	10.2 (5.7)	9.8 (6.8)	10.4 (5.1)	25.0 (8.6)	18.4 (7.2)	23.1 (8.1)
Average Pre-Test Score	65.8 (17.7)	64.8 (14.6)	54.9 (10.9)	56.9 (8.1)	96.1 (20.4)	99.0 (16.1)	88.5 (16.0)
Average Post-Test Score	82.1 (17.7)	83.6 (16.4)	73.7 (16.5)	76.8 (13.0)	110.5 (16.7)	112.7 (13.8)	104.1 (15.2)
Percentage of Classes in Schools Receiving Full Title I Funding for CR Program[b]	29	46	61	83	50	46	0
Percentage of Classes in Schools Receiving Full or Partial Title I Funding for CR Program[b]	75	63	87	88	72	65	0

a. Standard deviations are displayed in parentheses below the means.

b. Approximately 38 percent of the schools did not indicate the source of funds for their compensatory programs. The numbers on this line are based on the responses of those schools that indicated a source of funds.

c. See Note c, Table 4.

Table 6

AVERAGE CHARACTERISTICS BY CLASS TYPE FOR GRADE 6[a]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[c]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Average Enrollment	27.4 (6.0)	27.6 (5.5)	6.9 (4.1)	13.4 (11.2)	29.1 (4.7)	27.6 (5.2)	28.6 (6.6)
Average Number of Students Taking Tests by Type of Student	22.0 (9.7)	12.4 (7.8)	6.5 (3.5)	13.4 (9.7)	28.1 (7.8)	18.8 (8.2)	26.5 (7.2)
Average Pre-Test Score	83.3 (21.6)	86.1 (18.9)	66.5 (5.6)	74.3 (17.2)	111.4 (18.7)	117.1 (15.3)	112.5 (12.1)
Average Post-Test Score	93.6 (20.4)	96.1 (18.0)	80.3 (19.2)	85.2 (19.9)	118.4 (16.6)	122.9 (13.9)	118.0 (11.6)
Percentage of Classes in Schools Receiving Full Title I Funding for CR Program[b]	42	43	69	50	41	40	0
Percentage of Classes in Schools Receiving Full or Partial Title I Funding for CR Program[b]	79	68	88	67	58	65	0

a. Standard deviations are displayed in parentheses below the means.

b. Approximately 38 percent of the schools did not indicate the source of funds for their compensatory programs. The numbers on this line are based on the responses of those schools that indicated a source of funds.

c. See Note c, Table 4.

partial funding from ESEA Title I for the compensatory programs. Although the tables are basically self-explanatory, some additional discussion is necessary.

The average enrollment is based on an actual count of the number of students enrolled in each class in the analysis. The enrollment figure played an important part in calculating the average cost per student in the cost analysis. Smaller class sizes, while generally preferable for educational purposes, significantly raise the cost per student when the availability of resources is approximately equal. The average enrollment for the mixed classes of CR and NCR students is the sum of both the CR and NCR students. As described in the next paragraph, the split between CR and NCR can be estimated by examining the average number of students taking the tests. It should be pointed out again that the set of classes used to examine the CR students in mixed classes is slightly different from the set of classes used to examine the NCR students in similar classes. This produces the slight difference in average enrollment between the two groups.

The average number of students taking the tests is calculated as the sum of the number of students taking the pre-test and the number of students taking the post-test divided by two. All students were included in this calculation--not just those for whom both pre-test and post-test scores were available. A distinction was made between CR and NCR students when calculating this average for mixed classes. Thus, the total number of students taking at least one test in mixed classes can be approximated as the sum of the CR and NCR students taking at least one test. This was done to estimate the relative number of CR and NCR students in mixed classes.

The average number of students taking tests is generally less than the number of students enrolled in classes, probably because of absenteeism during the testing periods. But, in certain situations, the number of students taking tests is greater than the number enrolled. ETS considered the test results of any student who was ever enrolled in a class when calculating the characteristics of that class. If a student moved from one class to another during the school year, his test results would be included in both classes. Thus, classes that serve a variety of students during the year will be reported as having more students taking tests than are actually enrolled at any given time. ETS did not find that student movement was a particular problem. But it seems reasonable that special reading classes would have a slightly higher student turnover than conventional classrooms. This may explain the higher number of students taking tests in second-grade special reading groups.

However, this is probably not why the number of students taking tests in mixed classes is higher than it is for the corresponding enrollment. It should be remembered that the sets of mixed classes are not exactly identical when the CR and NCR students are considered separately. In particular, the mixed classes with small numbers of both CR or NCR students were eliminated. Thus, the average number of students taking tests in the remaining classes tends to overstate the true estimate.

ANALYSES OF VARIANCE

Tables 4, 5, and 6 presented the average characteristics for seven different types of classes for each grade. (Only six were presented for grade six because mixed special reading groups were eliminated.) It is of interest to determine if any statistically significant differences exist between class types in each grade. This will allow additional insight when interpreting the results of the cost-effectiveness analysis.

Analysis of variance (ANOVA) techniques were used to determine significant differences between any two types of classes in terms of enrollment, pre-test scores, and post-test scores. Because of the similarity between certain types of classes, all of the possible combinations could not be examined. However, a total of eight separate characteristic ANOVA calculations were made for each of the variables of interest in each grade. However, before presenting the results of these analyses, the procedure for interpreting their results must be discussed.

Two measures result from each ANOVA: the F test and the R^2 test. The F test is calculated as the ratio of the amount of variation between classes that can be explained by separating the classes into particular types to the amount of variation caused by experimental error. Obviously, as this ratio becomes greater, it is more likely that differences between group means are statistically significant and do not result from sample errors. At a critical value of this ratio, as determined from an F distribution and the degrees of freedom, the likelihood of the means not being significantly different becomes remote. In this study, the critical value is the ratio at which the differences between the means will be due to unexplained error only 5 percent of the time. While the exact value is dependent on the number of observations, the critical value of this ratio is approximately 3.0 for the typical sizes in this study. Above this value, the differences between the means are considered statistically significant.

The R^2 statistic, called the coefficient of determination, measures the strength of the differences between group means. It is calculated as the ratio of the unexplained variation to the total observed variation. The higher this ratio, the greater the likelihood of being able to predict the type of

class knowing only the value of the variable being examined. In this analysis, the R^2 statistic measures the proportion of the total variation that is explained by the class type. As an example, suppose that two types of classes have significant differences in mean pre-test scores, with type A having a higher average than type B. If the R^2 statistic between the two types of classes is high, say above 0.5, there is a high probability that the true class type can be identified for a randomly selected class based solely on an examination of the class pre-test score. If the class pre-test score is closer to the average for type A classes, it is very likely that it is a type A class. The reverse is true if the pre-test score is closer to the type B class. This statement can be made because there is not a great deal of overlap in the distribution of observed class pre-test scores between the class types.

On the other hand, if the R^2 statistic was low, say 0.1 as often happened in this study, the distribution of observed class pre-test scores for both class types is very overlapped. In that case, it would be unlikely that the true class type could be identified based solely on the pre-test score of a randomly chosen class.

There is a limitation in the interpretation of the ANOVAs that were conducted. These analyses used as their base the observed class average values of enrollment, pre-test scores, and post-test scores. Because class averages were used, additional within-class variance occurs that was not examined by the ANOVAs we performed. This problem of additional variance is aggravated by the variation in class size. However, this issue was not felt to be serious because of the general lack of difference between class types.

The first two ANOVAs, the results of which are presented in Tables 7 and 8, examine the differences in classes with CR students only compared with the CR students in mixed classes. This is done for both conventional classrooms and special reading groups.

In general, no differences are seen between CR students by themselves and CR students in classes with NCR students. This is true for class enrollment, pre-test scores, and post-test scores. The only exception is that the second-grade special reading groups are slightly smaller, on the average, when CR students are taught alone (8.5 versus 13.2). The division between the two types of classes is not sharp, with an R^2 of only 0.002. These two analyses seem to indicate that the effect of the two types of treatment (CR only versus CR/NCR mixed) is approximately equal.

Table 7

ANALYSIS OF VARIANCE BETWEEN CONVENTIONAL CLASSROOMS WITH CR STUDENTS ONLY AND CONVENTIONAL CLASSROOMS WITH A MIX OF CR AND NCR STUDENTS[a,b]

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	.82	.002	.00	.000	.006	.000
Pre-Test Scores	.01	.000	.37	.001	1.24	.004
Post-Test Scores	1.70	.005	.65	.002	1.45	.004

a. No significant differences exist at the 0.05 level.

b. See Note c, Table 4.

Table 8

ANALYSIS OF VARIANCE BETWEEN SPECIAL READING GROUPS WITH CR STUDENTS ONLY AND SPECIAL READING GROUPS WITH A MIX OF CR AND NCR STUDENTS

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	14.02[a]	.002	.86	.014	b	b
Pre-Test Scores	.03	.000	.01	.001	b	b
Post-Test Scores	1.32	.047	.63	.010	b	b

a. Significant differences exist at the 0.05 level.

b. The examination for sixth-grade mixed special reading groups was not conducted because of the small sample size.

However, significant differences can be seen between conventional classes and special reading groups, as demonstrated in the ANOVAs presented in Tables 9 and 10. In general, these differences exist for enrollment, pre-test scores, and post-test scores, regardless of whether the class is composed of CR students.

Examination of the characteristics data presented in Tables 4 to 6 shows that the enrollment for special reading groups is approximately one-third that of conventional classrooms. In addition, both the average pre- and post-test scores of the special reading groups are from 4 to 17 points below the corresponding scores for conventional classrooms. The exception to this generalization is the second grade. While the magnitude of the observed differences in the second grade is similar to that observed in the other grades, the differences are not statistically significant.

The next two ANOVAs, presented in Tables 11 and 12, examine the conventional classrooms with NCR students only. These classes are compared with the NCR students in the mixed CR/NCR classes and with the classes in NCR schools. While there are many exceptions, the following generalizations can be made:

- Classes with NCR students only are generally one or two students larger than the other two types.
- Classes in NCR schools generally have lower average pre- and post-test scores than classes with NCR students only.
- Classes with NCR students only generally have lower average pre- and post-test scores than NCR students in classes with a mix of CR/NCR students.

In the six ANOVAS discussed above, some statistically significant differences were observed. However, an examination of the R^2 statistic indicates that no strong differences exist between the various types of classes except when comparing special reading groups with conventional classes.

The last two ANOVAs, presented in Tables 13 and 14, examine the differences between CR and NCR students in both separate and mixed conventional classrooms. Significant differences are observed in these two comparisons. The enrollment in CR-only classes is slightly smaller, averaging one or two students fewer than the NCR-only classes. An examination of the enrollment in mixed classrooms was not conducted because the CR and NCR students were in the same classes. The average pre- and post-test scores for the CR students was always lower than the NCR students--usually by more than

20 points. In addition, the R^2 statistic was relatively high, particularly for CR and NCR students in mixed classes. This indicates very strong differences between the students in CR and NCR classes. Finally, the R^2 statistic describing the differences between CR and NCR classes decreases between the pre-test and post-test scores. (The only exception to this is the fourth-grade CR students only.) Thus, we see a more overlapped distribution for the post-test scores than for the pre-test scores.

Table 9

ANALYSIS OF VARIANCE BETWEEN CONVENTIONAL CLASSROOMS[b] AND SPECIAL READING GROUPS FOR CLASSES WITH CR STUDENTS ONLY

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	317.13[a]	.675	180.59[a]	.554	230.00[a]	.661
Pre-Test Scores	2.46	.015	13.39[a]	.077	13.00[a]	.001
Post-Test Scores	2.67	.016	6.73[a]	.040	7.80[a]	.062

a. Significant differences exist at the 0.05 level.

b. See Note c, Table 4.

Table 10

ANALYSIS OF VARIANCE BETWEEN CONVENTIONAL CLASSROOMS[b] AND SPECIAL READING GROUPS FOR CLASSES WITH A MIX OF CR AND NCR STUDENTS

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	165.22[a]	.397	224.32[a]	.480	c	c
Pre-Test Scores	3.56	.014	6.95[a]	.028	c	c
Post-Test Scores	1.47	.006	3.99[a]	.016	c	c

a. Significant differences exist at the 0.05 level.

b. See Note c, Table 4.

c. The examination of sixth-grade mixed special reading groups was not considered because of the small sample size.

Table 11

ANALYSIS OF VARIANCE BETWEEN CONVENTIONAL CLASSROOMS[b] WITH NCR STUDENTS ONLY AND CONVENTIONAL CLASSROOMS[b] WITH A MIX OF NCR AND CR STUDENTS

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	1.95	.006	9.67[a]	.025	7.73[a]	.002
Pre-Test Scores	7.08[a]	.020	1.95	.005	9.67[a]	.027
Post-Test Scores	18.72[a]	.023	1.53	.004	7.50[a]	.021

a. Significant differences exist at the 0.05 level.

b. See Note c, Table 4.

Table 12

ANALYSIS OF VARIANCE BETWEEN CLASSES WITH NCR STUDENTS IN SCHOOLS WITH CR PROGRAMS AND CLASSES WITH NCR STUDENTS IN SCHOOLS WITHOUT CR PROGRAMS[a]

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	4.14[b]	.029	5.07[b]	.036	.36	.002
Pre-Test Scores	14.50[b]	.034	16.18[b]	.036	.86	.002
Post-Test Scores	1.41	.033	14.91[a]	.033	1.83	.005

a. Only classes with NCR students only were considered.

b. Significant differences exist at the 0.05 level.

Table 13

ANALYSIS OF VARIANCE BETWEEN CONVENTIONAL CLASSES[b] WITH CR STUDENTS ONLY AND CONVENTIONAL CLASSES[b] WITH NCR STUDENTS ONLY

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Enrollment	4.36[a]	.020	8.29[a]	.038	5.72[a]	.025
Pre-Test Scores	84.77[a]	.281	132.45[a]	.386	108.93[a]	.328
Post-Test Scores	54.14[a]	.197	138.35[a]	.396	100.13[a]	.310

a. Significant differences exist at the 0.05 level.

b. See Note c, Table 4.

Table 14

ANALYSIS OF VARIANCE BETWEEN CR AND NCR STUDENTS IN MIXED CONVENTIONAL CLASSROOMS[b]

	Grade 2		Grade 4		Grade 6	
Variable Examined	F	R^2	F	R^2	F	R^2
Pre-Test Scores	495.96[a]	.504	614.50[a]	.546	336.49[a]	.453
Post-Test Scores	330.94[a]	.404	476.19[a]	.483	284.25[a]	.411

a. Significant differences exist at the 0.05 level.

b. See Note b, Table 7.

3

PROGRAM COST

The program cost data used in the cost-effectiveness analysis of compensatory reading programs were developed in the first phase of this study. A complete discussion of how these costs were calculated can be found in Volume I of this report. However, it is appropriate to summarize that work here.

COST-ESTIMATING MODEL

Most previous cost studies of education programs have relied on budget reports from the individual school a[illegible] primary data source for estimating costs. However, these budget-based estimates have proved inadequate for comparing alternative programs--primarily because the varied accounting procedures used in school districts make it difficult to compare costs. Further, such budgets generally do not highlight the concentrations of resources being examined. Finally, budgetary costs, because they indicate only the current year's expenses, neglect the value of resources inherited from previous years and distort the costs by including one-time capital expenditures.

The method adopted in this study, called the resource approach, was based on a rigorous assessment of the resources (i.e., staff, facilities, equipment, and materials) that were actually used in alternative compensatory reading programs. Specific attention was given to identifying where and how resources were concentrated by collecting all data at the classroom level. A total of 97 resources--including 39 types of staff members--were identified and handled separately in a computerized cost-estimating model.

To make the costs truly comparable, regional variations were eliminated by using average national values as the standardized resource cost (e.g., the salary of a teacher, including fringe benefits), and a proportional amount of the initial capital cost of acquiring the resource (e.g., the purchase price of a tape recorder). All resources used in a program were considered

and costed, even resources such as facilities that are generally available at no real cost to the school district. This was done to develop the total cost of the program, which is more comparable between school districts because of the variability in what is available at no cost to school districts.

The procedure used in the cost-estimating model involves estimating the quality of each resource available to a classroom. Using the standardized costs for each resource, the model calculates the total cost of employing that resource for reading. The total costs are then divided by the number of students in the class to yield an estimate for the average cost per student for each resource. Although the model examined 97 separate resources, it would have been difficult to analyze and interpret the results at this level of detail. Therefore, the costs were aggregated into 20 cost categories based on the function of the resources. An additional five cost categories were formed by aggregating the original 20 cost categories into higher levels. These 25 cost categories are listed in Table 15.

RESULTS OF COST ANALYSIS

An extensive discussion of the cost-analysis results was presented in Volume I of this report. The major results that affect the cost-effectiveness analysis are listed below:

- the average cost of all compensatory programs are about 50 percent higher than the cost of noncompensatory programs, and
- the costs of compensatory programs in ESEA Title I schools are 85 percent higher than noncompensatory programs.

Unfortunately, changes have occurred in the level and definition of the unit of analysis since the earlier phases of the cost-effectiveness analysis. Chapter 2 of this volume discussed these changes in detail. To briefly summarize them here, the original cost analysis used the grade average as the unit of analysis. The distinction between CR and NCR students was made using a principal-specified class label of CR or NCR. In contrast, the current analysis uses the class as the unit of analysis and attaches CR or NCR labels to the individual students in the class. In addition, the current analysis examines particular types of classes based on the following two criteria:

- whether or not the class is a special reading group, and
- its composition of CR and NCR students.

Table 15

COST VARIABLES

Description
1. Cost of Classroom Instructor
2. Cost of Teacher Aides
3. Cost of Reading Specialists
4. Cost of Other Special Teachers
5. Cost of Nonteaching Academic Staff
6. Cost of Nonteaching Non-Academic Staff
7. Cost of Supervision and Administration
8. Cost of Other Staff (Variables 5-7)
9. Cost of All Staff (Variables 1-7)
10. Cost of Facilities
11. Cost of Tape Recorders/Record Players
12. Cost of Teaching Machines
13. Cost of Computers
14. Cost of Movie/Filmstrip Projectors
15. Cost of Slide/Overhead/Opaque Projectors
16. Cost of Televisions/Video Tape Recorders
17. Cost of Other Equipment
18. Cost of All Equipment (Variables 11-17)
19. Cost of Basal Readers
20. Cost of Individual Reading Kits
21. Cost of Library Books
22. Cost of Other Supplies and Materials
23. Cost of All Materials (Variables 19-22)
24. Cost of District Centers
25. Cost of All Resources (Variables 9, 10, 18, 23, and 24)

While the previous cost analysis is still appropriate in its original context, particularly with respect to the impact of funding source, some additional analyses have been performed to provide a more meaningful basis for discussing program costs. The average cost in dollars per student for each of the various types of classes is presented in Tables 16, 17, and 18 for each grade. Because of the volume of these data, only the costs for selected resources are presented.

If these tables are compared with the results shown in Table 4-5 of Volume I, it can be seen that using various types of classes in the current study produced a marked decrease in the standard deviation of the costs associated with conventional classrooms. This is particularly true for conventional classrooms with CR students, which now have a standard deviation of one-half to one-third of that previously observed. This further demonstrates the usefulness of examining separately the various types of classes. This reduction in standard deviation is probably the result of separating reading groups from conventional classes. A corresponding increase in the standard deviation of the cost of such groups is observed, perhaps as a result of the high variability in enrollment relative to their size (see Tables 4, 5, and 6).

ANALYSES OF VARIANCE

As with the class characteristics in Chapter 2, it is of interest to determine whether there are any statistically significant differences in costs between the various classes. Again, ANOVA techniques were used to identify these differences. (Refer to Chapter 2 for a discussion of the interpretation of the results of such analyses.) For simplicity, the ANOVAs were conducted using only the total cost for all resources. If differences are not observed in total cost, it is unlikely that significant differences occur in the individual costs of particular major resources. Similarly, if significant differences in total cost do occur, examination of the various resource costs should indicate which resource costs are significantly different.

The first ANOVA, presented in Table 19, examines the differences in program cost across grades within class type. The only significant differences in cost occur for NCR students in classes by themselves. This is true for schools with or without CR programs. An examination of Tables 16, 17, and 18 reveals that the second-grade NCR-only classes are approximately 20 percent more expensive than either the fourth or sixth grades. The fourth-

Table 16

AVERAGE COST IN DOLLARS PER STUDENT FOR ALL CLASSROOMS FOR GRADE 2[a]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[b]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Class Instructor	73.77 (39.89)	62.12 (38.59)	64.47 (33.24)	75.94 (41.02)	73.49 (52.29)	64.53 (39.71)	76.91 (35.15)
Teacher Aide	20.32 (28.87)	7.52 (12.11)	41.04 (63.76)	29.96 (61.99)	6.50 (12.96)	6.61 (11.53)	11.46 (21.40)
Reading Specialists	31.05 (55.98)	14.78 (23.16)	320.11 (192.59)	214.47 (117.87)	10.09 (18.05)	15.90 (26.16)	14.52 (25.38)
Other Staff	31.25 (27.86)	17.56 (17.28)	41.01 (41.60)	50.93 (92.79)	15.41 (26.75)	17.56 (17.30)	16.59 (14.81)
All Staff	156.40 (87.12)	101.98 (53.94)	466.63 (236.80)	371.30 (213.37)	105.49 (71.72)	104.59 (55.09)	119.47 (56.65)
Facilities	37.69 (20.15)	29.94 (16.92)	168.81 (93.22)	139.31 (56.67)	34.88 (22.80)	31.49 (17.82)	37.65 (15.92)
Equipment	3.78 (8.85)	3.05 (2.94)	8.61 (7.10)	9.00 (7.42)	3.38 (4.65)	2.92 (2.92)	4.50 (2.94)
Materials	11.31 (15.29)	12.32 (4.69)	26.04 (13.97)	26.20 (8.33)	11.85 (5.90)	12.58 (5.72)	12.26 (7.18)
All Resources	215.51 (107.14)	150.66 (69.03)	676.19 (337.08)	552.09 (267.45)	158.40 (96.14)	154.53 (70.29)	176.18 (72.96)

a. Standard deviations are displayed in parentheses below the means.

b. As described in Chapter 2 and used in all analyses, conventional classrooms are considered as the typical classroom setting with one teacher teaching all subjects to 20 to 30 students.

Table 17

AVERAGE COST IN DOLLARS PER STUDENT FOR ALL CLASSROOMS FOR GRADE 4[a]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[b]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Class Instructor	66.37 (34.14)	61.16 (41.30)	64.57 (50.23)	79.52 (74.22)	53.36 (38.32)	58.82 (36.97)	57.95 (23.72)
Teacher Aide	16.68 (53.53)	7.42 (12.29)	28.69 (39.19)	25.52 (27.26)	3.45 (7.23)	6.92 (11.86)	7.18 (15.68)
Reading Specialists	26.95 (57.21)	15.46 (25.06)	273.09 (191.42)	240.04 (139.04)	10.10 (18.11)	15.90 (27.32)	9.65 (16.98)
Other Staff	22.90 (22.85)	18.80 (20.28)	44.25 (44.55)	59.88 (63.65)	18.60 (32.77)	18.04 (19.25)	14.34 (13.61)
All Staff	132.90 (94.49)	102.84 (57.60)	410.60 (195.13)	404.96 (181.72)	85.51 (52.35)	99.68 (53.89)	89.12 (36.17)
Facilities	32.47 (15.87)	29.10 (18.44)	159.13 (86.19)	143.69 (68.20)	26.24 (15.38)	27.95 (16.44)	28.65 (10.63)
Equipment	3.45 (4.01)	2.63 (2.29)	9.96 (8.81)	8.64 (7.28)	1.97 (1.93)	2.52 (2.24)	3.74 (3.21)
Materials	11.29 (16.10)	12.70 (5.46)	25.39 (12.73)	27.99 (17.33)	10.98 (5.31)	12.10 (5.49)	12.77 (6.43)
All Resources	185.79 (115.15)	150.50 (74.53)	610.15 (286.09)	588.85 (258.96)	128.48 (64.74)	145.13 (68.23)	137.20 (47.08)

a. Standard deviations are displayed in parentheses below the means.

b. See Note b, Table 16.

Table 18

AVERAGE COST IN DOLLARS PER STUDENT FOR ALL CLASSROOMS FOR GRADE 6[a]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[b]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Class Instructor	64.44 (40.57)	65.29 (51.66)	63.07 (32.27)	60.85 (22.94)	57.72 (36.63)	62.00 (47.14)	54.19 (26.64)
Teacher Aide	11.39 (22.86)	8.40 (33.46)	35.15 (45.04)	17.72 (22.04)	3.69 (10.20)	6.62 (21.76)	7.83 (17.09)
Reading Specialists	35.90 (62.89)	17.31 (27.96)	289.52 (239.10)	208.84 (83.21)	8.97 (15.22)	16.81 (26.83)	6.55 (8.97)
Other Staff	21.00 (21.82)	20.78 (23.81)	45.77 (47.75)	52.94 (61.38)	12.66 (12.63)	20.53 (24.10)	13.72 (9.81)
All Staff	132.73 (89.37)	111.78 (79.38)	433.51 (242.23)	340.34 (133.81)	83.04 (45.4)	106.06 (69.22)	82.28 (39.21)
Facilities	31.36 (18.31)	30.66 (20.97)	155.23 (83.42)	117.44 (38.80)	27.72 (15.55)	28.88 (18.78)	25.20 (10.09)
Equipment	3.00 (3.84)	2.80 (2.93)	10.01 (7.53)	9.56 (6.66)	1.98 (1.78)	2.73 (2.69)	2.86 (1.87)
Materials	12.00 (17.93)	11.57 (5.28)	25.08 (8.88)	28.08 (12.56)	10.53 (4.54)	11.83 (5.75)	12.09 (8.50)
All Resources	183.74 (114.37)	159.40 (98.72)	630.44 (331.25)	495.98 (170.18)	127.13 (62.08)	151.78 (86.98)	125.94 (52.48)

a. Standard deviations are displayed in parentheses below the means.

b. See Note b, Table 16.

and sixth-grade classes are approximately equal in cost. In addition, most of the difference in cost can be attributed to differences in staff costs. This result is in complete agreement with the previous cost analysis presented in Volume I (see Table 4-10 of Volume I).

Table 19

ANALYSIS OF VARIANCE OF TOTAL COST BETWEEN GRADES 2, 4, AND 6

	Compensatory Students				Noncompensatory Students		
	Conventional[b] Classrooms		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
F	3.12	.77	.38	.12	5.38[a]	1.16	10.11[a]
R^2	.017	.002	.009	.005	.035	.003	.119

a. Significant differences exist at the 0.05 level.

b. See Note b, Table 16.

Table 20 presents the results of eight separate ANOVAs that examine various combinations of classes. The first three ANOVAs compare conventional classrooms with CR students only versus conventional classrooms with a mix of CR/NCR students, conventional classes with NCR students only, and conventional classes in NCR schools. Generally, conventional classrooms with CR students only are significantly more expensive than other types of conventional classrooms, ranging from 15 to 45 percent more expensive. The only exception is the significant difference in cost between sixth-grade classes with CR students only and those with a mix of CR and NCR students.

The next two ANOVAs examine the costs of conventional classrooms that have NCR students only compared with conventional classrooms that have a mix of CR and NCR students and conventional classrooms in NCR schools. Only the fourth-grade classrooms with a mix of CR and NCR students proved significantly more expensive than the NCR-only classrooms. From Table 17, it can be estimated that such classes cost approximately 17 percent more. This general conclusion is in agreement with the conclusions reached in the original cost analysis (see Volume I, page 95).

Table 20

ANALYSES OF VARIANCE FOR TOTAL COST

	Grade 2		Grade 4		Grade 6	
	F	R^2	F	R^2	F	R^2
1. Conventional classrooms with CR students only versus conventional classrooms with a mix of CR and NCR students[b,c]	48.59[a]	.120	11.92[a]	.033	3.50	.012
2. Conventional classrooms with CR students only versus conventional classrooms with NCR students only	16.43[a]	.069	17.73[a]	.078	22.53[a]	.092
3. Conventional classrooms with CR students only versus conventional classrooms in NCR schools	6.22[a]	.032	9.20[a]	.049	9.60[a]	.066
4. Conventional classrooms with NCR students only versus conventional classrooms with a mix of NCR and CR students[d]	.17	.000	4.12[a]	.001	8.23	.024
5. Conventional classrooms with NCR students only versus conventional classrooms in NCR schools	1.39	.010	.76	.005	.03	.000
6. Special reading groups with CR students only versus conventional classrooms with CR students only	177.13[a]	.521	182.02[a]	.529	118.40[a]	.501
7. Special reading groups with CR students only versus special reading groups with a mix of CR and NCR students	2.40	.040	.09	.001	e	e
8. CR students in mixed classes versus NCR students in mixed classes	.038	.002	.72	.001	.69	.002

a. Significant differences exist at the 0.05 level.

b. See Note b, Table 16.

c. The cost for CR students in mixed conventional classrooms was used.

d. The cost for NCR students in mixed conventional classrooms was used.

e. The examination for sixth-grade mixed special reading groups was not conducted because of the small sample size.

ANOVAs 6 and 7 in Table 20 compare the costs of special reading groups having CR students only versus conventional classrooms with CR students only and special reading groups that have a mix of CR and NCR students. Special reading groups with CR students only are significantly more expensive than conventional classrooms with CR students. In general, such groups are two to three times more expensive. In addition, the R^2 statistic is very high. Fifty percent of the variation in cost between the two types of classes can be explained by the class type. This indicates a strong difference in cost between the two classes. At least part of this difference in cost can be attributed to the fact that students in special reading groups receive instruction from two separate teachers. When the cost of special reading groups that have CR students only is compared with special reading groups that have a mix of CR and NCR students, no significant cost differences are found.

The last ANOVA compares the cost of providing instruction to CR students in mixed conventional classrooms with the cost for NCR students in the same type of class. As discussed in Chapter 2, a slight difference in the two sets of mixed classes is caused by the elimination of small classes. The last ANOVA was conducted to determine if this slight difference produced significant differences in cost. As can be seen in Table 20, there are no diffferences.

TYPAL ANALYSIS

Because of the problems associated with collinearity between various categories of cost, as well as a hypothesis that a certain mix of resources may appear more effective than other mixes, we did a simple typal analysis of the cost variables. Sometimes called elementary factor analysis[1] or cluster analysis, this technique used the intercorrelation matrix of the variables under study (various categories of cost in this case). Basically, the method seeks to establish clusters of variables, with the clusters determined by the criterion that each variable in a given cluster must show a higher correlation with some other variable in the cluster than with any variable outside the cluster. The object is to group variables into clusters

1. Louis L. McQuitty, "Elementary Factor Analysis," Psychological Reports, 9 (1961), 71-78.

that are relatively tightly knit within and relatively unrelated (i.e., uncorrelated) with other clusters.

Mechanically, the process by which this is done is straightforward. The correlation matrix is calculated, and the highest value in it is found. The two variables that account for this correlation (let us call them x and y) are the basis for the first cluster. Then, the correlation of each with other variables is examined to determine whether some third variable (say, w) should be brought into the cluster on the basis of being more highly correlated with x or y than with any other variable in the matrix. If w is brought in, its correlations are similarly examined to see if yet further variables can be brought into the first cluster. The process is reiterated from the beginning, excluding those variables from the matrix already in the one cluster, and is continued until the matrix is exhausted.

A hypothetical example will indicate the logic and simplicity of this analysis. Suppose that six cost variables had the following correlation matrix:

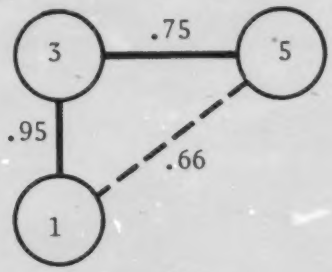
Correlations Between Variables

	Variable					
	1	2	3	4	5	6
1		.10	.95	.05	.66	.20
2	.10		.15	.80	.02	.71
3	.95	.15		.25	.75	.20
4	.05	.80	.25		.22	.91
5	.66	.02	.75	.22		.15
6	.20	.71	.20	.91	.15	

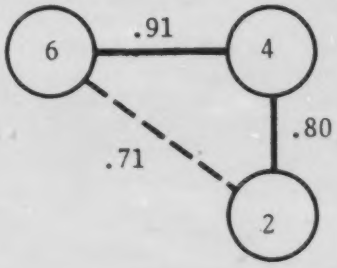
Inspection shows that r_{13} is the largest correlation in the matrix, and hence variables 1 and 3 form the beginning of the first cluster. Then, we ask if further variables can be brought into this cluster based on their correlation with 1. It can easily be seen that, in all cases, the correlation of 1 with other variables is smaller than some other correlation of the variable. Thus, variable 1 brings no others into the cluster. However, when we examine variable 3, we see that its correlation with 5 is higher than any other correlation of 5, so 3 brings 5 into the cluster.

If we reiterate this process, deleting variables 1, 3, and 5 from the matrix, the largest remaining correlation is seen to be $r_{46} = 0.91$, so variables 4 and 6 form the beginning of a second cluster. Following the procedure outlined above shows that variable 2 is then brought into the second cluster through its link with 4. The analysis is now complete, and the six original cost variables are unequivocally identified with one of the two clusters. Visually we might review the results as:

Cluster 1:



Cluster 2:



We see that the technique has very clearly partitioned the original variable set into subsets ("clusters") that are fairly highly intercorrelated and seemingly independent (the highest correlation <u>across</u> clusters is $r_{34} = 0.25$).

Typal analysis has been shown to give results that are very similar to factor analysis in cases that involve clusters of the type discussed: tight relationships with groups and little cross-cluster correlation. When significant cross-cluster linkage does exist, care must be taken in interpreting the results to ensure that the method does not arbitrarily partition groups on the basis of very small differences. This may be avoided by examining all such linkages. If cross-cluster linkage exists, clusters can be combined on the basis of the relationships between, within, and across the cluster correlations.

This technique was applied to a correlation matrix of all cost categories in each of the following three strata:

- second-grade compensatory classes,
- second-grade noncompensatory classes in schools with compensatory programs, and
- second-grade groups in schools with no compensatory programs.

The analysis was restricted to these three strata on the premise that, if the results proved promising in our attempts to identify cost-effectiveness relationships, we would continue analyzing the remaining strata.

The purpose of this analysis was to construct categories of resources with costs that appeared to move up or down together. This would reduce the number of independent variables in regression equations. At the same time, it would greatly reduce the problem of multicollinearity in the independent variables.

The concept seemed sound, and the results were somewhat consistent across strata. Figures 1, 2, and 3 present the typal analysis results showing the first cluster produced for each stratum. The three variables--cost of classroom instructor, cost of facilities, and cost of other supplies and materials--were always the first variables assigned to cluster 1 for each stratum. The method also assigned two additional variables to cluster 1 for compensatory groups. Seven remaining clusters were similar across strata to some extent, but have been omitted for brevity. Although the results of the analysis could constitute trends of resource allocation, it was soon realized that some of them were related to the cost model calculations and their inherent assumptions. This is particularly true of facilities and classroom instructor costs since there was no variation in facilities cost except the proportion of time used that was also applied to classroom instructor cost. This does not necessarily invalidate the results, but it certainly ties them to assumptions made concerning the cost model calculations.

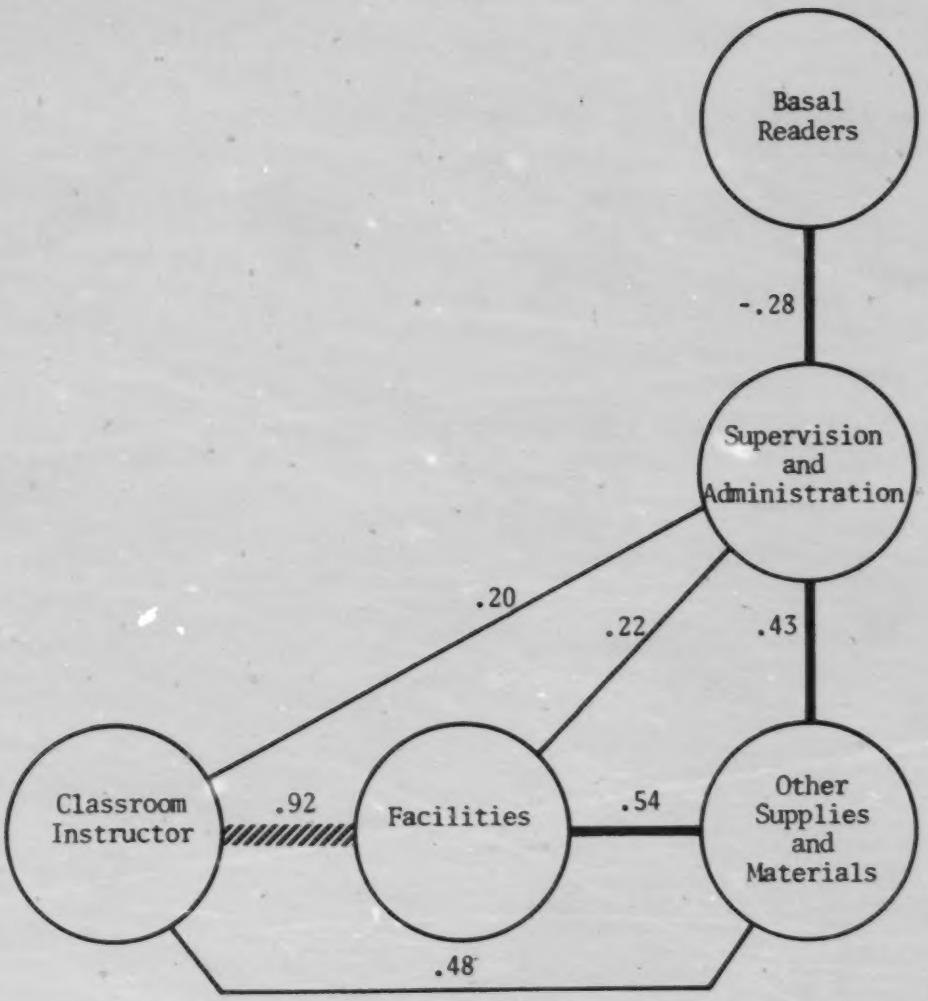


Figure 1: TYPAL ANALYSIS RESULTS--SECOND-GRADE COMPENSATORY GROUPS

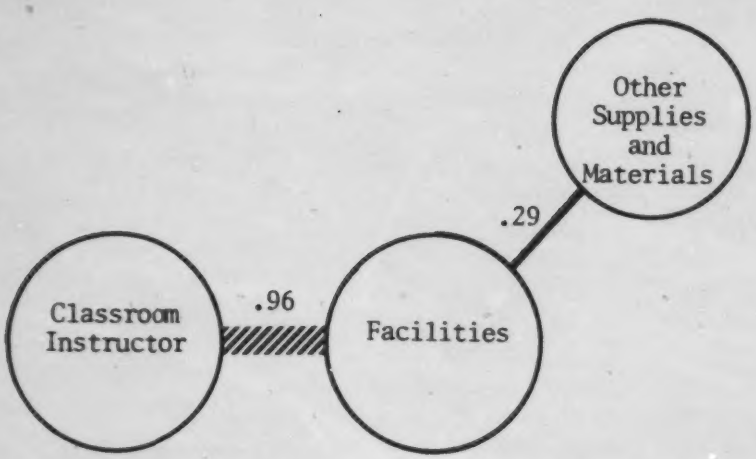


Figure 2: TYPAL ANALYSIS RESULTS--SECOND-GRADE NONCOMPENSATORY GROUPS IN SCHOOLS WITH COMPENSATORY PROGRAMS--CLUSTER 1

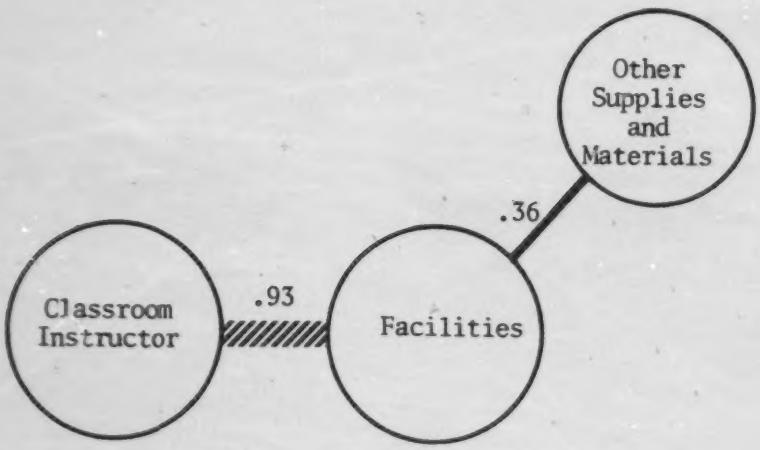


Figure 3: TYPAL ANALYSIS RESULTS--SECOND-GRADE GROUPS IN SCHOOLS WITH NO COMPENSATORY PROGRAMS--CLUSTER 1

4

PROGRAM EFFECTIVENESS

As pointed out in the Introduction, the data collection and analysis activities involved in this cost-effectiveness study of compensatory reading programs was a cooperative effort between ETS and RMC. RMC was responsible for estimating the cost of the various programs, while ETS was responsible for estimating their effectiveness. Chapter 3 described the cost estimates developed by RMC. This chapter will present and discuss the measure of effectiveness developed by ETS.

MEASURE OF EFFECTIVENESS

A meaningful measure of effectiveness for use in a cost-effectiveness study must not only identify which programs are more effective, but must also indicate how much more effective these programs are. The measure of effectiveness chosen for this study by ETS was the residual gain score. The residual gain score is the difference in the actual average post-test raw score for a class and the raw score that the students would be expected to achieve considering their initial skill levels and experience. In algebraic form, the measure of effectiveness is calculated as follows:

$$\text{effectiveness} = \text{actual post-test} - \text{expected post-test.} \qquad \text{Eq. 1}$$

It might have been possible to calculate a simple measure of effectiveness by subtracting the actual pre-test score from the actual post-test score. However, such a procedure has an inherent error that precludes its use. A high score on an achievement test usually indicates that an individual has achieved a higher level of skill than another individual who has scored lower. However, the high skill level represented by one more correct answer does not always represent the same magnitude of increased skill. This is caused by a nonlinearity between what might be called the "true" achievement level and

the test score. This nonlinearity makes a simple subtraction of post-test and pre-test scores meaningless as a measure of effectiveness. The use of the residual gain scores is an attempt to compensate for this nonlinearity.

The concept of residual gain score is best demonstrated by an example. Assume that four classes--called A, B, C, and D--are being examined. All four classes are essentially identical in their initial conditions of achievement, experience, and environment. The assumed results of a post-test given to each of these classes is presented in Table 21. Further, assume that a separate analysis has shown that classes with similar initial conditions should be expected to have a score of 60 on the post-test. The difference between the actual post-test score and a score of 60 is the measure of effectiveness. This measure of effectiveness is calculated for each class in our example, and is shown in Table 21.

Table 21

RESULTS OF HYPOTHETICAL STUDY

Class	Actual Post-Test Score	Measure of Effectiveness[a]
A	80	20
B	70	10
C	60	0
D	50	-10

a. Calculated using equation 1 with an assumed expected post-test score of 60.

Examination of Table 21 shows that the measure of effectiveness is proportional to the actual post-test score. Class A scored 10 points higher than Class B on the post-test, and its measure of effectiveness is also 10 points higher. In this sense, there is no difference between the actual post-test score and the measure of effectiveness. What is different between the two is that the measure of effectiveness has a zero point that divides effective and ineffective classes. Thus, classes A and B are effective, while class D is ineffective. This is indicated by the algebraic sign of the measure of effectiveness. While it is tempting to say that class A is twice as effective as class B, the use of such a ratio scale is inappropriate. If such a scale were used, class A would be infinitely more effective than

class C (the result of dividing a measure of effectiveness of 20 by zero). The only appropriate statement that can be made about the relationship in effectiveness between classes involves an interval scale. Thus, class A is 10 units more effective than class B; and class B is also 10 units more effective than class C.

The theoretical advantage of the residual gain score is that it allows meaningful comparisons to be made at different levels of initial conditions. It should be remembered that the hypothetical example given above assumed that the four classes all had the same initial conditions. If all classes examined in this study had the same initial conditions, the actual post-test score would be an appropriate measure of effectiveness. But, this is not the case because initial conditions vary widely. Classes with higher initial levels of achievement should be expected to score higher on the post-test. If the actual post-test score was used, the class with the higher pre-test score would more than likely be called the more effective class because it would probably have the higher post-test score. However, such a procedure is not really measuring program effectiveness. Post-test scores are influenced far more by pre-test scores, and thus initial conditions, than by the effect of a particular program. Thus, if actual post-test scores are used as a basis of comparison for classes with widely different pre-test scores, the observed differences are more likely the result of differences in initial conditions rather than program effectiveness. The use of the residual gain score is an attempt to measure program effectiveness independent of the initial conditions. Such a measure will indicate which classes are effective and which are ineffective, using an interval scale similar to that demonstrated in Table 21.

The interpretation of the residual gain score as a measure of program effectiveness is relatively straightforward. The major methodological difficulty in its use is the calculation of the expected post-test score, or the score that would be expected considering the initial conditions. As indicated in the example above, the expected post-test score for a particular class is a function of the initial conditions of achievement, experience, and environment observed in that class. A class with an average achievement lower than another class would be expected to score lower on the post-test, assuming all else is equal. Similarly, a class that has had richer prior experiences would be expected to score higher on the post-test than another class that did not have those experiences.

The usual method for calculating the expected post-test score is through regression analysis. This technique involves a statistical procedure for

removing the observed variation in the post-test scores for a large set of classes caused by initial achievement levels (as measured by a pre-test) and prior experience and environment (as measured by various socio-economic indicators). The socio-economic indicators were originally considered as covariates along with pre-test scores in the calculation of the expected post-test scores performed by ETS. However, as is often true in educational studies, the various socio-economic indicators were found to be highly correlated with the pre-test scores. The use of highly correlated variables in a regression analysis produces meaningless, extraneous results. For this reason, socio-economic variables were eliminated from the calculations conducted to determine the expected post-test scores. Only the pre-test scores were used. It should be pointed out that eliminating the socio-economic variables as covariates does not remove their effect from the ultimate measure of program effectiveness, even if they are correlated with pre-test scores. These variables still remain part of the overall treatment to which students are subjected. Pre-test scores measure only those effects prior to the treatment. This does not necessarily invalidate the use of residual gain scores as a measure of effectiveness. However, their accurate use requires assuming that the nonprogram effects are constant across all classes.

CALCULATION OF THE MEASURE OF EFFECTIVENESS

The general procedure followed by ETS in calculating the expected post-test score involved estimating a relationship between pre-test and post-test raw scores to minimize the variance (i.e., the square of the difference between actual post-test and expected post-test for any class). Two functional forms were used to estimate this relationship: linear and quadratic. The linear functional relationship takes the following algebraic form:

$$\text{expected post-test} = a_1 + a_2\ (\text{pre-test}) + \varepsilon, \qquad \text{Eq. 2}$$

where a_i = equation coefficients and ε = error term.

The quadratic functional relationships take the following algebraic form:

$$\text{expected post-test} = a_1 + a_2\ (\text{pre-test}) + a_3\ (\text{pre-test}) + \varepsilon. \qquad \text{Eq. 3}$$

The actual technique used a statistical procedure for estimating the equation coefficient, a_i, to minimize the square of the error term.

Once the expected post-test score for any particular class is calculated using either equation 2 or 3, the measure of effectiveness for the class can be calculated through the use of equation 1. Both the linear and quadratic

forms will be discussed in greater detail in later sections of this chapter.

TESTS USED FOR PRE-TEST AND POST-TEST SCORES

In the sample schools, all children in grades 2, 4, and 6 took a battery of tests in the fall and in the spring. The tests included a reading achievement measure composed of two subtests of the grade-appropriate Metropolitan Achievement Test and either the Cooperative Primary Test or the Step Reading Test. The children's attitudes toward reading were also measured based on ETS-developed tests. One test was given for children in the second-grade and another for fourth- and sixth-grade children. These attitude tests were administered orally to eliminate any bias that might be introduced to the test because a child was unable to read well enough to understand it.

The reading achievement tests were chosen to measure a wide range of achievement since a great deal was not known about the student population being tested. The Metropolitan Achievement Test was felt to have a "low floor" and the Step and Cooperative Primary Tests to have adequately "high ceilings." Different forms were given in the fall and spring to compensate for expected gains in the children's reading achievement. The reading achievement test took a maximum of 1-1/2 hours to administer, while the attitude test took 10 to 15 minutes to administer.

The reading achievement test battery consisted of 127 items for the second grade, 145 items for the fourth grade, and 155 items for the sixth grade. The attitude test consisted of 4 practice items and 24 real items for the second grade and 4 practice items and 45 real items for the fourth and sixth grades. Table 22 presents the breakdown of the tests and subtests administered and the number of items in each.

LINEAR FORM OF EFFECTIVENESS MEASURE

Initially, ETS developed a series of linear equations to predict an expected post-test score for each of the grades involved in the study. The functional form of this equation was presented in equation 2 above. The equation coefficients for each of the grades can be found in Table 23. At that time, ETS felt that it was not necessary to distinguish between the various types of classes.

Table 22

TESTS USED

Test or Subtest	Fall Form	Spring Form	No. of Items
2nd Grade			
MAT Sentences			13
MAT Stories			29
MAT Reading (Sentences and Stories)			42
MAT Word Knowledge			35
MAT (Primary Level) Total (Reading and Word Knowledge)	G	F	77
Cooperative Primary Reading	12A	12B	50
Battery Total			127
4th Grade			
MAT Reading			45
MAT Word Knowledge			50
MAT (Elementary Level) Total	G	F	95
Cooperative Primary Reading	23A	23B	50
Battery Total			145
6th Grade			
MAT Reading			45
MAT Word Knowledge			50
MAT (Elementary Level) Total	G	F	95
Step II (Level 4) Reading	A	B	60
Battery Total			155

Table 23

EQUATION COEFFICIENTS FOR LINEAR FORM

[expected post-test = $a_1 + a_2$ (pre-test)]

	Coefficients	
	a_1	a_2
Grade 2	47.2042	.7110
Grade 4	25.5775	.8804
Grade 6	21.2474	.8639

These equations were used to make an estimate of the expected post-test score that each class in the study should have achieved based on its pre-test score. Based on this expected post-test score, and the actual observed post-test score, a measure of the effectiveness for each class was calculated using equation 1.

The measure of effectiveness then formed the basis for a cost-effectiveness analysis similar to the one presented in Chapters 5 and 6 of this report. However, the results of this analysis were weak and inconsistent. Therefore, a separate examination was conducted by RMC into the effectiveness measure and its relationship to other variables. A scatter diagram of post-test scores versus pre-test scores revealed that a nonlinear relationship existed between post-test and pre-test scores. The plot for all classes in grade 2 is presented in Figure 4. The equation for the expected post-test scores is also plotted.

An examination of this plot shows the apparent tendency for the expected post-test scores to be above the general grouping of observed post-test scores for classes with either low or high pre-test scores. Similarly, the expected post-test scores are generally below the observed post-test scores for classes with pre-test scores that are in the middle. Because the measure of effectiveness is calculated as the difference between expected and observed post-test scores, there is an apparent tendency to identify more effective programs in classes with middle pre-test scores. This tendency was confirmed through an examination of the mean values of effectiveness for various sets of the classes in grade 2, partitioned by pre-test scores. Since this tendency results from the definition of the measure of effectiveness, it is an inherent weakness in the measure.

For this examination, the classes in grade 2 were partitioned into five sets, depending on the class average pre-test scores. The pre-test break points were scores of 53, 65, 76, and 88. These scores divided the classes into groups of approximately equal size. The mean value of effectiveness for each partition was calculated and is presented in Table 24. As you can see, the average effectiveness measure for classes with a pre-test score of less than 53 or greater than 88 is equal to approximately -0.28. This indicates that these classes are less effective than average. Similarly, the average for all other classes is approximately 3.0. This supports the

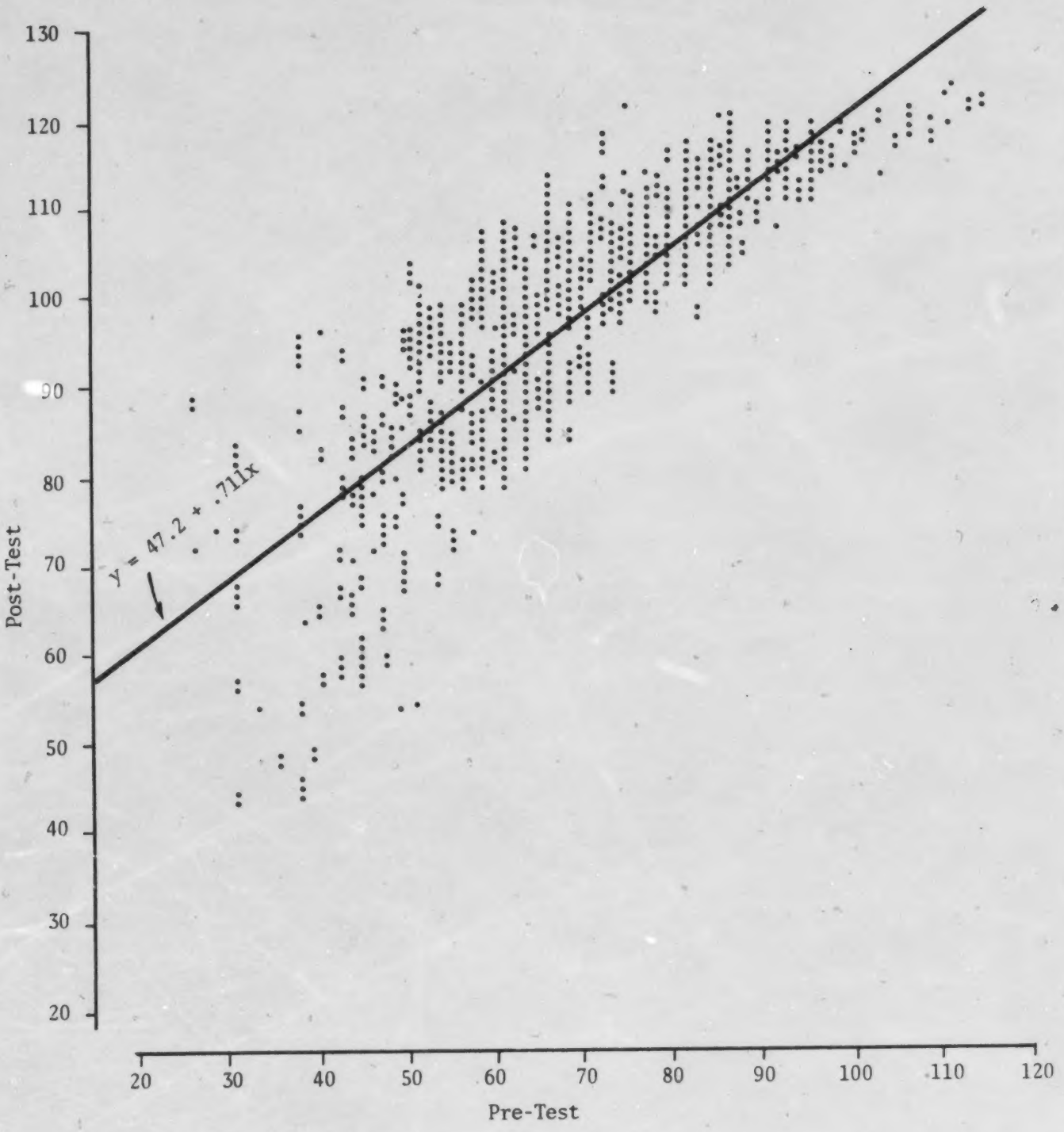


Figure 4: POST-TEST VERSUS PRE-TEST RESULTS--GRADE 2

tendency observed in Figure 4. In addition, an analysis of variance was conducted to determine if the observed differences were statistically significant. The results of this ANOVA, also presented in Table 24, indicate that the differences are significant at the 0.05 level.

Table 24

ANALYSIS OF VARIANCE OF THE MEASURE OF EFFECTIVENESS USING PARTITIONED PRE-TEST SCORES FOR ALL CLASSES IN GRADE 2

Partition of Pre-Test Scores	Mean
Pre-Test< 53	-4.664
53 ≤ Pre-Test< 65	2.601
65 ≤ Pre-Test< 76	3.480
76 ≤ Pre-Test< 88	2.922
Pre-Test≥ 88	- .999
F test on differences between partitions = 23.02[a]	
R^2 ratio between partitions = 0.12	

a. Significant differences exist at the 0.05 level.

The exact reasons for the nonlinear relationships are not completely known, although they are probably the result of the ceiling effect of the tests given. Since the pre- and post-tests were essentially the same, students scoring high on the pre-test were unable to properly indicate their increased achievement during the school year because they could not answer that many more questions correctly. Whatever the reason, the fact that the measure of effectiveness varies with the pre-test scores violates the theoretical basis of the residual gain scores. In particular, the expected post-test scores are lower than the observed post-test scores for all but a limited number of classes at the extremes of the test.

If these measures of effectiveness were used, additional complications would affect the analysis. For example, Chapter 2 showed that CR students in both conventional classrooms and special reading groups had lower pre-test scores than average. Thus, there would be a strong tendency for such classes to have low or negative measures of effectiveness for the reasons

demonstrated in Table 24. Furthermore, Chapter 3 showed that the same classes were likely to cost more than average. This would produce a negative relationship between cost and effectiveness (i.e., negative effectiveness associated with high cost). This conclusion seemed unwarranted considering the observed limitations in the linear model of effectiveness. Therefore, no additional analyses using the linear relationship were conducted.

QUADRATIC FORM OF EFFECTIVENESS

At RMC's request, ETS redeveloped equations to estimate the expected post-test scores using the quadratic functional form presented in equation 3. Because of the lack of parallelism of the regression curves, separate estimating equations were developed for each type of class in each grade. The coefficients for each type of class can be found in Table 25. These equations were used to make an estimate of the expected post-test score and its associated measure of effectiveness.

For each type of class being examined, the plot of the observed post-test versus pre-test scores was constructed. On top of this distribution, the curve presenting the expected post-test score was then plotted. A series of figures similar to Figure 4 resulted. These figures have been included as Appendix A of this volume. An example of one of these curves is presented in Figure 5. Here, the pre- and post-test results for second-grade students in mixed classes are shown, with both linear and quadratic equations superimposed. The expected post-test scores resulting from the quadratic form more accurately represent the average post-test scores for the overall range of observed pre-test scores.

To confirm this impression, an examination was conducted of the mean values of effectiveness for various sets of all the classes in grade 2, partitioned by pre-test score. The procedure for this examination was identical to that presented in Table 24. The results are shown in Table 26, and indicate that, while there was still a slight tendency to underestimate the effectiveness of classes with both low and high pre-test scores, no statistical significance could be attached to this difference.

Table 25

EQUATION COEFFICIENTS FOR QUADRATIC FORM
[expected post-test = a + b (pre-test) + c (pre-test)^2]

	Compensatory Reading Classes			Noncompensatory Reading Classes in a School That Has Compensatory Reading			Compensatory Reading Students in Mixed Classes			Noncompensatory Reading Students in Mixed Classes			Classes in Noncompensatory Schools		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
Grade 2	32.9179	1.0593	-.0020	-24.0161	2.5749	-.0116	-11.8896	2.4776	-.0121	15.6284	1.7884	-.0077	12.3210	1.7697	-.0073
Grade 4	-2.3362	1.6336	-.0048	21.5616	1.1530	-.0022	-1.5869	1.6476	-.0049	21.4513	1.0506	-.0013	-16.2577	1.9148	-.0061
Grade 6	3.5063	1.2783	-.0023	18.1445	.9347	-.0003	24.4760	.8102	.0002	29.1572	.7266	.0006	-47.9829	2.0654	-.0052

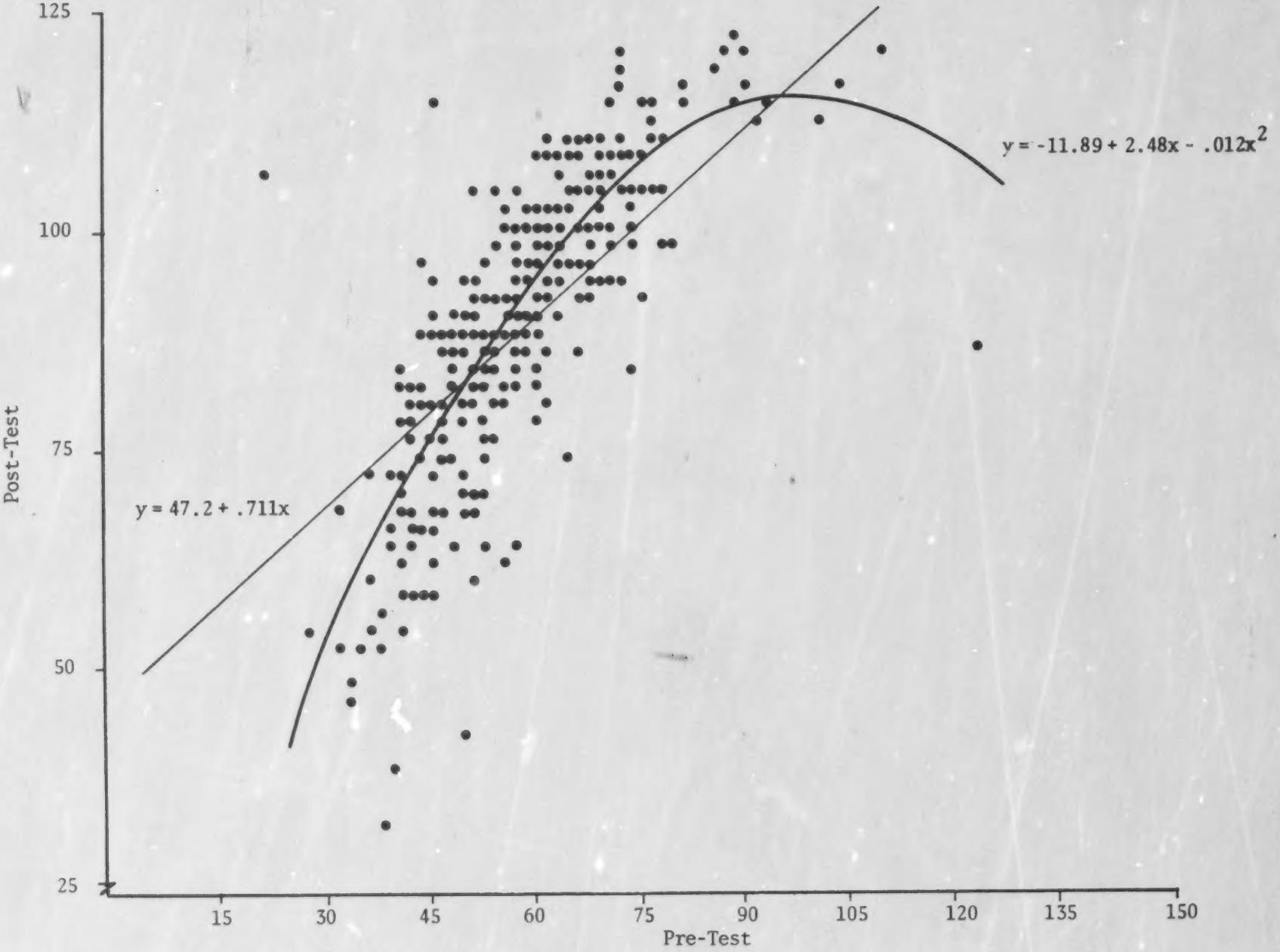


Figure 5: PRE-TEST AND POST-TEST RESULTS--CR STUDENTS IN MIXED CLASSES--GRADE 2

Table 26

ANALYSIS OF VARIANCE OF THE QUADRATIC MEASURE OF EFFECTIVENESS USING PARTITIONED PRE-TEST SCORES FOR ALL CLASSES IN GRADE 2

Partition of Pre-Test Score	Mean
Pre-Test< 53	-.7299
53 ≤ Pre-Test< 65	+.8682
65 ≤ Pre-Test< 76	+.8414
76 ≤ Pre-Test< 88	-.5613
Pre-Test≥ 88	-.4184
F test between partitions = 1.466[a]	
R^2 ratio between partitions = .0063	

a. No significant differences exist at the 0.05 level.

LIMITATIONS OF THE MEASURE OF EFFECTIVENESS

Two important factors are embedded in the measure of effectiveness developed for this analysis that limit the use and interpretation of the analysis results. The first factor is the discrepancy in the sets of students taking the pre-test and post-test. The second involves the problem of heteroscedasticity, or an undesirable relationship between the variance of the measure of effectivenss and the pre-test score. It is important to fully understand the limitations imposed by these two factors. Therefore, each will be discussed below.

Different Students Taking Pre-Tests and Post-Tests

When calculating the average pre-test and post-test scores necessary for determining the measure of effectiveness, ETS included the test results for any students who were ever enrolled in the class. This was done even if the student did not take both the pre- and post-tests. This lack of congruence between the two sets of students taking the test is of concern because it could introduce variance that is not the result of a particular educational approach. This additional variance is independent of the test score, and thus is not factored out.

The theoretical basis for using the residual gain score as a measure of effectiveness is the elimination of all variation associated with the initial conditions (as measured by pre-test). The remaining variation is solely due to the effect of the educational treatment received in the class, plus a small random error. This random error can be attributed to either testing error or absenteeism. The reading tests administered as part of this study are not a completely accurate way of measuring true student achievement. The testing conditions and the attitude of the student on the day of the test influence the actual test score. Under a different situation, the test result might be different. These conditions are the result of random events that both hinder and aid different students. A student missing either the pre-test or the post-test because he was absent on the day the test was given introduces additional error and uncertainty into the evaluation. As long as the student was enrolled in the class during most of the year, the variation is random. It is hoped that, in the current analysis, both these errors are random and small.

If, however, a student did not take both tests because he either enrolled in the school after the pre-test or left the school before the post-test was administered, there is a possibility for serious error. First, if post-tests for such students are used, nontreatment effects may influence the post-test because the students were not enrolled in that class for the full program. Second is a potential for serious bias. To demonstrate this, let us postulate a class with a program that is effective, but that has a high turnover during the year. The high turnover rate will reduce the likelihood that the students who have been helped are still in the class. In addition, because the program is thought to be effective, the replacements (who did not take the pre-test) would more than likely be the students with lower achievement. The net result will be that the post-test scores are significantly lower than would be expected (i.e., the higher scoring students influenced by the program have been replaced by lower scoring students not influenced by the program). This represents a significant bias.

The exact magnitude of this problem is not known, although some insight is provided in the histogram presented in Figure 6. There, the frequency of the differences between the number of pre-tests and post-tests is plotted for second-grade conventional classrooms with CR students only. It can be seen that the number of students taking the pre-test is no more than two

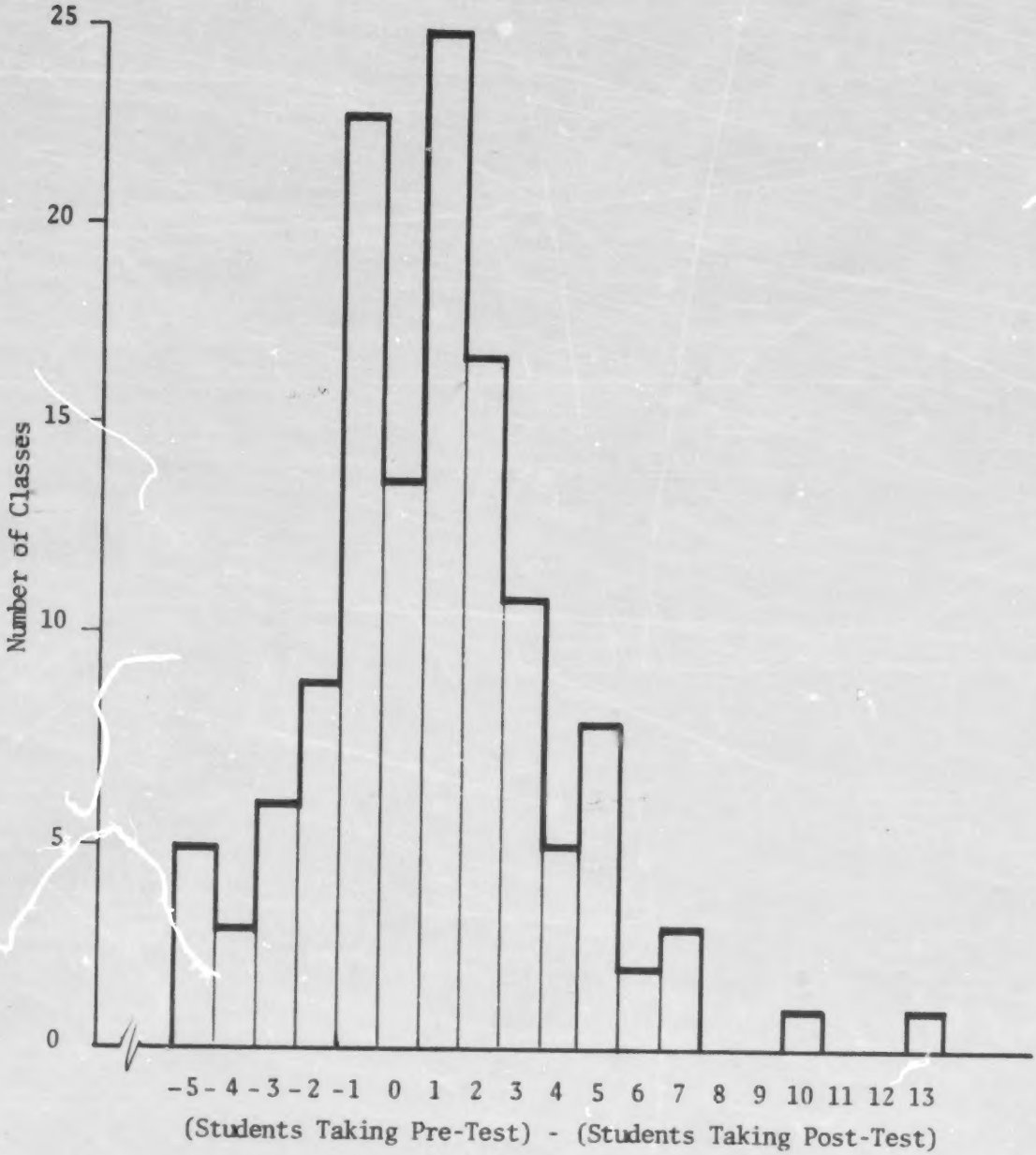


Figure 6: HISTOGRAM OF THE DIFFERENCES IN NUMBER OF STUDENTS TAKING THE PRE-TEST AND POST-TEST IN SECOND-GRADE CONVENTIONAL CLASSROOMS FOR CR STUDENTS ONLY

different from the number of students taking the post-test for approximately 66 percent of the class. There is a slight tendency, however, for more students to take the pre-test than the post-test, with an overall average of 0.9 more students taking the pre-test. The standard deviation is 2.95.

The histogram is only a minimal demonstration of the problem because no attempt was made to identify the actual number of students who took both tests. While the problem is not significant in the majority of cases, it is quite large for particular classes. Thus, this problem represents a limitation in interpreting the results.

Heteroscedasticity

The measure of effectiveness, as measured by the residual gain score, is calculated through regression techniques. The measure of effectiveness is, in fact, generally called the residual in a regression equation. Certain statistical properties of the residual of a regression equation must be true to ensure statistical acceptance. The three most important are: (1) the mean value of the residuals should be near zero, (2) the residuals should be normally distributed, and (3) the residuals should not be a function of any independent variable.

With regard to the residual gain scores used as the measure of effectiveness in this analysis, the validity of the first two requirements can be demonstrated in Figure 7. Here, the frequency of each value of the measure of effectiveness is plotted for all classes in the second grade with CR students. The mean value of this distribution, equal to 0.9, is sufficiently near zero considering the range (-30 to +36) so as not to invalidate the residuals. Likewise, an examination of the distribution itself indicates that it approximates a normal distribution. Similar distributions were developed for other types of classes, all of which demonstrated reasonable normality.

However, the validity of the third requirement for acceptance of the residuals is suspect. Figure 8 presents a plot of the observed values of effectiveness as a function of the pre-test score. An obvious relationship between the measure of effectiveness (as measured by the residual gain score) and the pre-test score can be seen. As indicated by the dashed lines, the envelope of observed values of effectiveness decreases with increases in pre-test scores. This same relationship can be observed in Figure 4. The

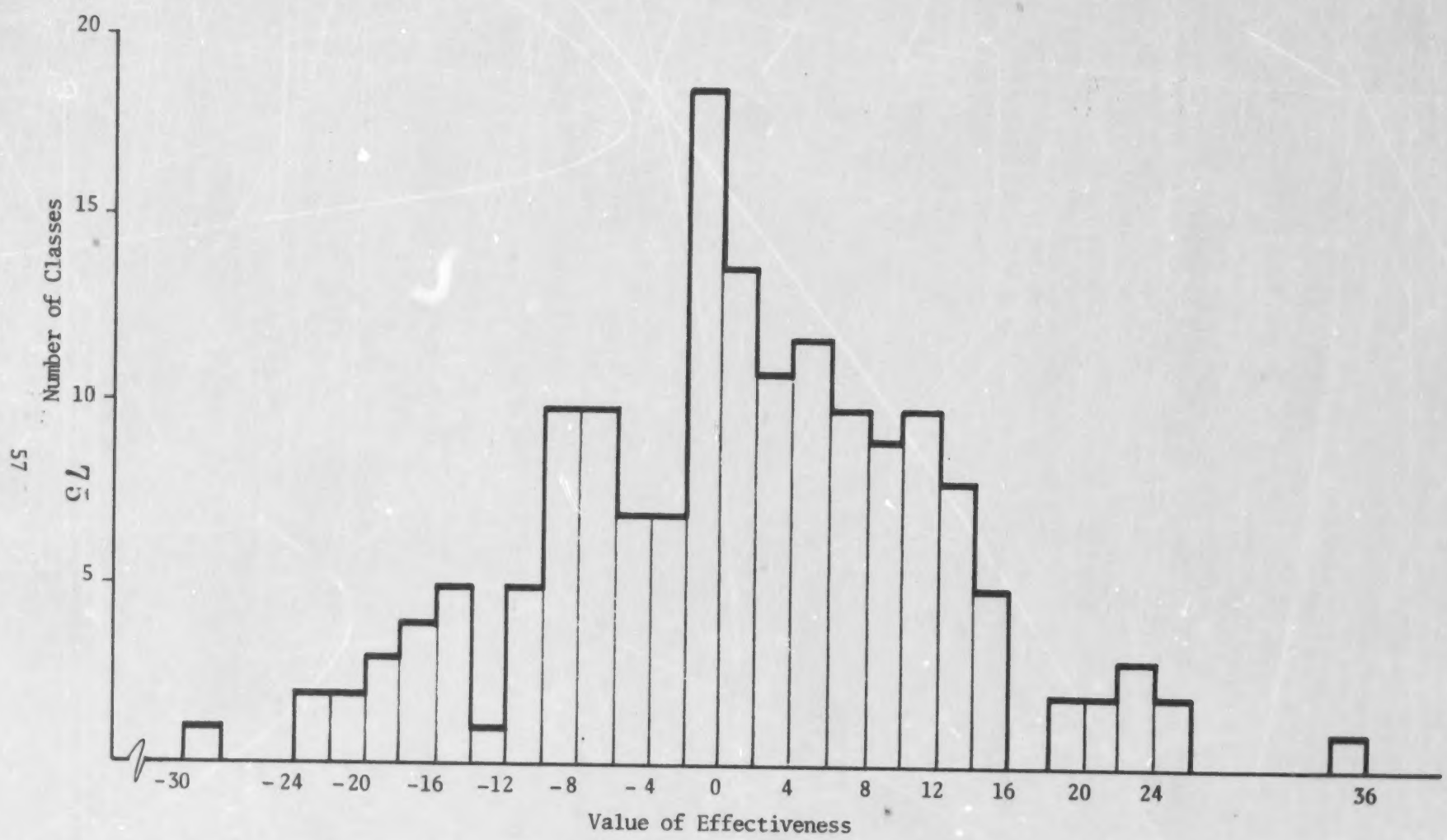


Figure 7: DISTRIBUTION OF EFFECTIVENESS FOR SECOND-GRADE COMPENSATORY GROUPS

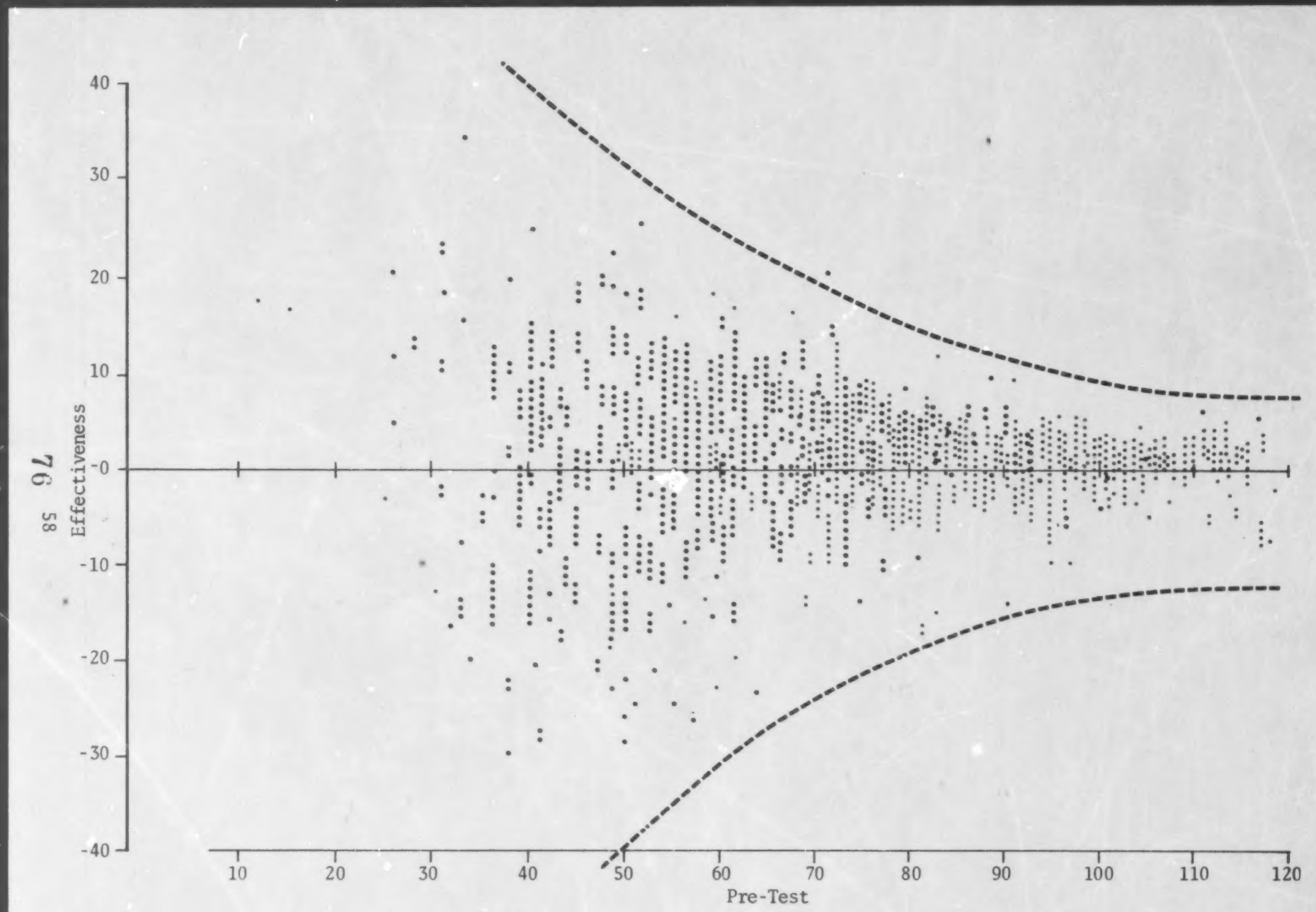


Figure 8: QUADRATIC EFFECTIVENESS VERSUS PRE-TEST SCORE--GRADE 2

relationship cannot be eliminated in the original regression analysis because it is embedded in the tests and the variance between the tests, and is not directly associated with initial conditions. This relationship creates a problem, called heteroscedasticity, because it raises statistical questions about the validity of the original equations used to estimate the expected value of the post-test score.

The most likely cause of this relationship between effectiveness and pre-test scores is a decrease in the sensitivity of the test with increases in the test scores. In other words, it is more difficult to score one more correct answer when you are near the top of the test than when you are in the middle or low end. This fact is demonstrated by the flattening out of the curves in Figures 4 and 5 near their top ends.

Besides the statistical problems caused by heteroscedasticity, other problems are present that involve how heteroscedasticity affects the interpretation of the measure of effectiveness. An examination of Figure 8 shows that the best class scoring 50 on the pre-test had a measure of effectiveness of approximately 25. However, the best class scoring 100 on the pre-test had a measure of only about 5. If the two measures of effectiveness were used alone, the interpretation would be made that the class with the lower pre-test score was 20 units more effective than the class with the higher pre-test score. This appears unreasonable considering both classes were the best in their pre-test groups. This is not to say that the lower scoring class was not more effective, just that it is difficult to believe it was that much more effective. Thus, the essence of this problem is the apparent difficulty of comparing the measure of effectiveness for different groups of classes with widely varying pre-test scores.

One possibility of overcoming this problem is to assume that the most effective classes for each pre-test level are equally effective. A transformation could then be made to the measure of effectiveness that eliminates the problem of heteroscedasticity. The effect of this type of transformation is demonstrated in Figure 9. The variance of effectiveness is now uniform with the pre-test score. Such a transformation has been developed by Parks for econometric models.[1] Time did not allow us to derive such a

1. R. E. Parks, "Estimation with Heteroscedastic Error Terms," Econometrica (October 1966).

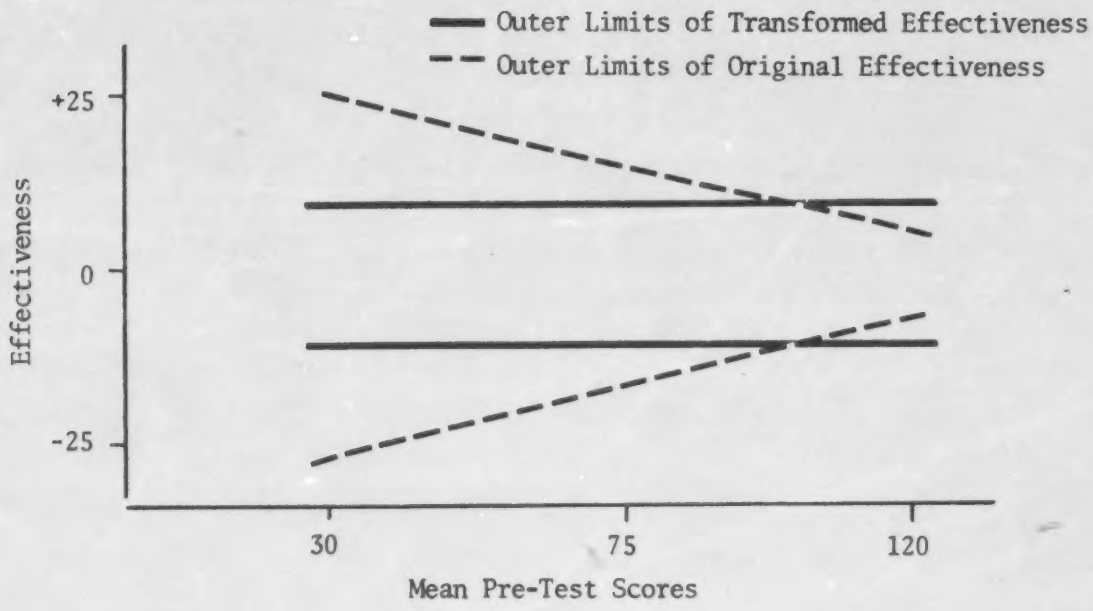


Figure 9: EFFECT OF TRANSFORMATION TO CORRECT HETEROSCEDASTICITY

transformation. However, a crude transformation, as suggested by Draper and Smith, was tried.[1] The following form was used to recalculate the measure of effectiveness for second-grade conventional classrooms with CR students only:

new effectiveness = (old effectiveness) x (pre-test score) x (0.01).

A small analysis was conducted that was similar to the one described in the next two chapters. The results of this analysis were compared with the results obtained using the original measure of effectiveness. The results were quite similar with respect to all relevant statistics. Because of this, no additional transformations were made.

The use of a transformation as described above may, in fact, create more problems than it solves. The goal of the residual gain score calculation used as a measure of effectiveness is to eliminate all variance other than what can be attributed to the effect of educational treatment. Unfortunately, the error associated with the ability of the test to actually measure achievement cannot be eliminated. The variance in test results that can be attributed to treatment effects is less for classes with high pre-test scores because of the decreased sensitivity of the tests at the higher end. It is possible, though not likely, that the entire variance in test scores observed at the high end is the result of measurement error associated with the test rather than any treatment. Under any circumstances, the ratio between the variance attributed to measurement of error and the total variance caused by both treatment and measurement error will increase for classes with higher pre-test scores. Since the transformation described above increases the size of the variance with increasing pre-test score, both the treatment effect and the measurement error increase with increasing pre-test score. Thus, while such a transformation may solve the problem of heteroscedasticity, it may produce measurement errors that are no longer small--a factor that tends to cloud any analysis. This entire area should be investigated before transformations are made.

PRESENTATION OF MEASURES OF EFFECTIVENESS

Despite the limitations discussed above, it was decided to continue the analysis using the measure of effectiveness.

---

1. Draper and Smith, Applied Regression Analysis (New York: Wiley, 1968).

Table 27 presents the mean measure of effectiveness and its standard deviation for each type of class using the quadratic equations developed in Table 25. Because of the differences in the equations used to calculate this measure for the different class types, it is not really appropriate to compare the means for the different class types. It should be pointed out that the procedure used to calculate the measure of effectiveness (i.e., residual gain score) causes the average measure of effectiveness to approach zero. Thus, no significance should be attached to the differences in mean effectiveness even though they are not all zero. None are significantly different from zero.

It is interesting to observe that the standard deviation of the measure of effectiveness has a tendency to be inversely proportional to the mean value of the class pre-test score as determined from Tables 4, 5, and 6. In general, the special reading groups that have the lowest pre-test scores have the highest standard deviations of effectiveness. Conversely, the NCR classes, which have the highest pre-test scores, have the lowest standard deviations. This observation confirms the existence of the heteroscedastic problems.

Table 27

AVERAGE CLASS EFFECTIVENESS[a,b]

	Compensatory Students				Noncompensatory Students		
	Conventional Classrooms[c]		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
Grade 2	1.302 (10.826)	.865 (7.680)	-.136 (11.035)	1.742 (8.059)	.336 (4.623)	.051 (4.466)	.181 (5.649)
Grade 4	-.871 (5.790)	.052 (6.527)	1.398 (7.617)	1.217 (5.683)	-.611 (4.179)	.308 (4.706)	.244 (4.429)
Grade 6	.657 (5.341)	.283 (6.540	2.662 (10.130)	-.255 (7.83)	-.016 (4.242)	.337 (4.201)	.181 (4.123)

a. Standard deviations are displayed in parentheses below the means.

b. This measure of effectiveness is the quadratic measure of effectiveness developed in equation 3, using the coefficients presented in Table 25.

c. See Note b, Table 16.

5

MEASUREMENT OF COST EFFECTIVENESS

The primary objective of the entire study was to determine the relative merits of alternative compensatory reading programs. One approach toward accomplishing this goal was to examine the relative effectiveness of such alternative programs. This approach was also used by ETS in its analysis. A second approach, which is the subject of this report, is to conduct a cost-effectiveness evaluation of the various programs. Such an evaluation attempts to answer the following two questions:

- If functional relationships exist between the cost of program resources and student achievement gain, what is the nature of these relationships?
- Are certain compensatory reading programs more cost effective (i.e., they offer a greater return in achievement gain for each dollar spent) than others?

The previous chapters have discussed the three types of data required to accomplish the goals outlined above. These data include program characteristics, program cost, and program effectiveness. This chapter addresses cost effectiveness, how it can be measured, and how it can be used. In essence, then, this chapter presents the basic methodology for evaluating the cost effectiveness of alternative compensatory programs. The results of applying this methodology are presented in Chapter 6. Before beginning a discussion of the more technical aspects of measuring cost effectiveness, it seems appropriate to define what is meant by the term cost effectiveness.

COST EFFECTIVENESS

In its most elementary form, the measurement of cost effectiveness is an attempt to measure the economic efficiency of a process. Often there are alternative means of achieving a particular end to the process. If this end requirement is fully understood and well defined, the measurement of

cost effectiveness reduces to determining which alternative is the least costly. In the vast majority of situations, however, the desired end is not well defined nor fully understood. When that occurs, a cost-effectiveness analysis must examine the relationships between cost and effectiveness to provide insight and guidance into how to efficiently allocate resources in achieving various goals. This is the purpose of the analysis presented in this report.

Until very recently, public school administrators have been willing to accept the idea that the cost of a program is synonymous with the value of the output resulting from the program. In essence, they felt that the more money spent on a program, the more successful the program. Even though the benefits or outputs of education were not clearly known or capable of being valued, they were nonetheless considered worth the cost required. Because of the many intangibles involved in the education process, it was felt that the school was not a factory that could be run on the "profit" motive.

These arguments can never be completely refuted. The school must provide a minimum level of education to everyone--regardless of the cost. However, the economic issues represented by the cost of the program cannot be ignored. Educational funds provided at the local, state, and federal levels are all under close and critical scrutiny from other competing social services. Therefore, the general benefit to society of educational funds--as measured by the effectiveness of the educational programs--must be demonstrated if educators are to expect continued funding.

This is not to say that a cost-effectiveness analysis is easy or that it will be completely acceptable to either the educator or other social agencies. The meaning and measurement of effectiveness are particularly difficult because it is here that the intangibles of the educational process come to the surface. The previous chapters examined the measures of effectiveness that will be used in this analysis. Many problems and some serious limitations were found. Nevertheless, it is important to attempt to evaluate the cost effectiveness of educational programs.

The primary beneficiaries of this analysis will be educational administrators and planners. Many alternative education programs are available, and all of them have the goal of quality education. However, each alternative has a different approach and different resource requirements. This analysis examines the relationships between the resource requirements and

the effectiveness of the alternative programs in light of various student characteristics. One goal of the analysis is to provide educational administrators with greater insights for efficiently allocating scarce resources to provide effective educational programs. It is no longer sufficient to say "more is better."

COMPARISON OF COSTS ACROSS CLUSTERS

The methodology, as presented in the original proposal, was altered considerably as a result of substantial changes in the overall study plan. In fact, a cost-effectiveness evaluation of the various program clusters was never formally undertaken. After clustering classrooms (the basic unit of analysis) into 14 groups based on demographic and program characteristics, ETS analyzed the effectiveness of these clusters, but found no differences in effectiveness. Therefore, RMC's first step in the cost-effectiveness analysis was a simple matter of comparing costs across the 14 clusters and identifying those project types that required significantly higher or lower costs than the average project. This analysis was conducted during early 1974 and was presented in Volume I of this report. The basic results of that analysis revealed some statistically significant differences between the costs of the various clusters. Additional analysis showed that clusters 3 and 5, averaging approximately $350 per student, appeared more expensive than the rest of the clusters, which averaged in the neighborhood of $200 per student. A complete description of the various clusters can be found in the ETS report.[1]

EXPLANATION OF OBSERVED CHANGES IN EFFECTIVENESS IN TERMS OF PROGRAM COSTS

The original analysis plan suggested a multiple linear regression analysis as the appropriate way to identify how program effects (achievement gain) relate to the program cost categories. The plan suggested the inclusion of program characteristics with program costs in the regression analysis to estimate equations of the following form:

$$PE = A + b_1 F + b_2 E + b_3 S + b_4 M + \ldots b_i X_i \ldots + \text{error},$$

where

PE = program effectiveness,

---

1. Rubin, et al., A Descriptive and Analytical Study of Compensatory Reading Programs.

F = facility costs,

E = equipment costs,

S = staff costs,

M = material costs, and

X_i = other parameters used to explain variance of effectiveness. The "X" parameter identifies all program characteristics and demographic variables other than the costs that affect program effectiveness.

This plan, although highly appropriate in theory, carries with it many assumptions and restrictions that have, up to this point, never been thoroughly tested in an educational context. For example, the variables used in calculating the estimate should be relatively independent of each other if spurious variation in their regression coefficients is to be eliminated. The typal analysis in Chapter 2 showed some correlation between the various cost categories. Additionally, the correlation between program costs and program characteristics is not fully understood.

A good example of this problem is demonstrated by the cost analysis results in Volume I of this report. There, it was shown that compensatory classes receiving funding under ESEA Title I cost between 35 and 85 percent more than classes that receive no federal assistance. Since eligibility for ESEA Title I generally requires certain economic and demographic conditions, some correlation between cost and these types of conditions will certainly be present. Work performed by ETS indicated that pre-test scores were highly correlated with many of these same socio-economic conditions.

A second example of possible problems concerns the true relationships between effectiveness, resource utilization (sometimes availability), and resource costs. Economic theory is based on assumptions that relationships between cost and effectiveness are mathematically functional. However, this may not be true with respect to resource costs, resource utilization, and educational achievement as calculated in this study. Traditionally, strong functional relationships have been proven between input and output variables before undertaking the study of cost-effectiveness relationships. This is hardly the case in the field of education, and the ways in which costs have been grouped for this study have never been significantly related (in terms of utilization) to achievement gain.

Here, it is important to point out that the planned regression approach was completed. However, many other types of analyses were required to

Another reason for studying the correlation coefficients is to become aware of the highly collinear independent variables in the regression equations. Generally, where two independent variables are more highly correlated to each other than to the dependent variable, the effects of collinearity will give spurious results in the regression coefficients. Thus, care must be taken not to exclude highly collinear variables in a regression equation. This turns out to be a significant problem in the data. As shown in the typal analysis in Chapter 2, many of the cost variables are highly correlated, while the correlations between the various cost variables and program effectiveness are generally weak.

Regression Equations

A cost-effectiveness analysis using multivariate regression techniques is an effort to find significant relationships between certain resources (measured by their cost) and effectiveness. It should be remembered that cost is only an indicator of the utilization by a class of a certain resource based on the availability of that resource to the class. Of course, a resource such as a class instructor can be used in many different ways, depending on the personality of the teacher, his educational background and training, and his previous experience in teaching. Cost gives the relative economic value of different resources to the success of the program.

If one extrapolates economic theory to educational theory, the coefficients of the independent variables in the regression equations could conceivably be interpreted as the marginal return to effectiveness of a dollar spent on that resource. This is a very complex extrapolation, and should probably be avoided in this study. The reader will notice that the values of the coefficients from the regression equations are rarely presented. The concept that a dollar spent on equipment returns a slightly larger increase in effectiveness than a dollar spent on a teacher is not necessarily sound. A teacher highly influences how instructional equipment is used, and, of course, instruction probably would not take place at all if a teacher were not present in the classroom.

The value of the regression analysis lies in determining what proportion of the variability in effectiveness can be explained by cost variables, which cost variables are statistically significant in explaining a portion of this variation, and their associated sign. The sign is important because it

validate the regression results. These other analyses were performed to test assumptions required by the regression approach and to support our conclusions resulting from the regression analysis.

ANALYTICAL APPROACH

Despite the limitations discussed above and in Chapter 4, a concerted effort was made to examine all meaningful relationships between cost and effectiveness. The following approaches were used:

(1) an examination of correlation coefficients,

(2) the identification of relationships through linear regression techniques,

(3) an analysis of variance examination of the costs of effective programs,

(4) an independent analysis of post-test scores at various cost levels,

(5) a simple comparison of class types, and

(6) a separate examination of the costs of unusually effective programs.

Each approach is explained below, along with its strengths and weaknesses.

Correlation Coefficients

The correlation coefficient, called R, is a measure of how closely the relationship between two variables approaches a linear relationship. The coefficient, which can range from -1 to +1, is useful in a cost-effectiveness analysis as a means of determining the correlation between cost and program effectiveness. If the R calculated between cost and effectiveness is greater than zero, higher costs are generally associated with higher levels of effectiveness. Conversely, if the R is less than zero, higher costs are generally associated with lower levels of effectiveness. The closer R is to either -1 or +1, the stronger the relationship between cost and effectiveness. When R is zero, there is no relationship between cost and effectiveness. However, the correlation coefficient does not necessarily indicate any cause and effect relationship between cost and effectiveness. It only indicates that certain conditions occur together (e.g., high cost and high effectiveness). Only variables that significantly correlate with effectiveness have a reasonable probability of proving significant in the linear regressions examined in the next approach.

indicates whether the resource is positively or negatively related to effectiveness. A positive relationship indicates that higher quantities of the resource are associated with higher levels of effectiveness.

As described earlier, linear regression techniques employ the following generalized equation:

$$PE = C + \Sigma a_i X_i + \varepsilon,$$

where

PE = program effectiveness,

X_i = program variables such as the cost for a particular resource or other program characteristic,

C, a_i = regression coefficients, and

ε = error term.

The techniques are employed to estimate the regression coefficient to minimize the error term. Since a linear function form is used, the derivative of the equation taken with respect to a particular independent variable, say X_i, results in a value that can be considered as the incremental increase in effectiveness resulting for one more unit of the independent variable. For example, if the independent variable was the cost of the class instructor, the derivative would indicate the incremental increase in effectiveness that would be realized from using the equivalent of a one dollar increase in instructor services (say, by spending slightly more time teaching reading, or by using a better qualified, but more expensive, teacher). Using the form of the equation given above, the incremental increase in effectiveness can be easily determined as:

$$\frac{\partial PE}{\partial X_i} = a_i,$$

or simply the regression coefficient of that variable. If the coefficient for one variable is higher than it is for another variable, this result can be interpreted as meaning the first variable is more cost effective than the second variable because it has a higher incremental increase in effectiveness when an additional dollar is spent for both.

As a result of the uncertainty associated with the measures of effectiveness, we will not stress the importance of the size of these coefficients in this report. Of more interest, however, is the sign of the regression coefficient. If this sign is positive, increases in the availability and use of this resource in a classroom would appear to increase the effectiveness of the instruction in that class. Conversely, a negative sign indicates

that an increasing quantity of the resource tends to produce a reduction in effectiveness.

Analysis of Variance of Effective Programs

The proper use of multiple linear regression techniques requires a great deal of statistical integrity that, as we have already pointed out, may be lacking in the data. In particular, the problems of the nonlinearity and heteroscedasticity in the estimation of effectiveness may make the results of the linear regression analysis meaningless.

In an attempt to overcome these problems, an alternative approach to evaluating the cost effectiveness of various resources was tried that did not depend as heavily on the measure of effectiveness. This approach involved stratifying the individual classrooms into separate cells by pre-test and post-test scores. The classrooms were stratified by pre-test score to allow an examination of the impact of the pre-test score on cost effectiveness. The stratification by post-test score was accomplished by subjectively dividing the post-test scores for each pre-test stratum into the post-test scores that could be considered better than average, those that were about average, and those that were below average. The average post-test grouping was determined from the expected post-test score used in the effectiveness calculation. Classrooms whose post-test scores were within a reasonable range from the expected scores, usually ± 30 percent, were considered average. Classrooms with post-test scores above this range were considered better than average, while classrooms with post-test scores below this range were considered ineffective. This stratification scheme, which eliminated the interval scale basis of the measure of effectiveness, reduced the problems of nonlinearity and heteroscedasticity.

Once the strata were defined, analysis of variance techniques were performed to determine if significant differences existed in the costs associated with each of the cells. When it was determined that the cost of a resource in an effective cell was statistically greater than the corresponding average or ineffective cells, this resource was identified as having a positive relationship between its availability and the effectiveness of the class. If the reverse was true, a negative relationship was identified. A more detailed discussion of this approach and its results are presented in Chapter 6.

Independent Analysis of Post-Test Scores

The regression equation approach attempted to examine the relationships between cost and effectiveness. The analysis of variance approach attempted to overcome some of the statistical problems embedded in the measure of effectiveness by examining the differences in costs between classes that seemed to be effective and classes that seemed to be ineffective. As an alternative approach, it was decided to examine the educational gains associated with various levels of cost.

This approach again involved estimating the expected post-test score as a function of the pre-test score using equation 3 in Chapter 4. You will recall that this equation was used in calculating the quadratic measure of effectiveness as shown below:

$$\text{expected post-test} = a_1 + a_2\ (\text{pre-test}) + a_3\ (\text{pre-test})^2 + \varepsilon,$$

where

a_i = equation coefficients, and
ε = error term.

In the original calculation of the measure of effectiveness, all classes of a particular type were used when the regression equation was solved. This time, however, the classes were divided into three, approximately equal subsets that represented high, medium, and low total costs. The expected post-test regression equation was solved separately for each subset. The resulting curves of expected post-test scores for high, medium, and low costs were then plotted against pre-test scores. These curves were examined to determine which cost levels were associated with higher expected post-test scores at each pre-test score level.

This approach allows a form of cost-effectiveness analysis that is sensitive to the pre-test level and the type of class being examined. In addition, the problem of heteroscedasticity, though not eliminated, is minimized. What is being examined is the estimated post-test score--not a regression error term as in the calculation of the measure of effectiveness.

Simple Comparison

RMC undertook a systematic review of the various characteristics, cost, and effectiveness data presented in Chapters 2, 3, and 4. This was done in an effort to uncover any additional statements that might be made about the relationships between cost and program effectiveness.

Separate Examination of Unusually Effective Programs

As part of its evaluation, ETS conducted in-depth observations of 29 schools. These schools were selected to represent a wide range of effective and ineffective programs. Neither the school being observed nor the observers knew its effectiveness rating. Thus, this effort took the form of a double blind experiment.[1] A total of five schools were identified as being unusually effective based on both the test results and the observational data. Analysis of variance techniques were used to identify significant differences in costs between classes in these five schools and other classes included in the study.

1. Trismen, A Descriptive and Analytical Study of Compensatory Reading Programs: Final Report, Volume II, Educational Testing Service (April 1976).

6

RESULTS OF THE COST-EFFECTIVENESS ANALYSIS

This chapter presents the results of RMC's analysis of the cost effectiveness of alternative compensatory reading programs. This cost-effectiveness analysis represents an attempt to examine the relationships between the characteristics, cost, and effectiveness of various reading programs. The following six approaches were used to examine these relationships:

(1) an examination of correlation coefficients,

(2) the identification of relationships through linear regression techniques,

(3) an analysis of variance examination of the cost of effective programs,

(4) an independent analysis of post-test scores at various cost levels,

(5) a simple comparison of class types, and

(6) a separate examination of the cost of unusually effective programs.

Although a detailed discussion of these methodologies can be found in Chapter 5, the results of each approach are presented below.

The results of the cost-effectiveness analysis presented in this chapter are quite extensive. Nevertheless, neither time nor sufficient money was available to examine all possible combinations of cost and effectiveness in our search for relationships. Still, every effort was made to examine characteristic combinations that we felt were indicative of particular situations. However, it is possible that other combinations should have been examined. This may be particularly true for special reading classes, which generally were not examined in great detail. Therefore, the reader should bear this in mind when reviewing the results in this chapter.

CORRELATION COEFFICIENTS

A correlation coefficient was calculated between total cost and pre-test scores, post-test scores, and effectiveness for each type of class in the analysis. These coefficients are presented in Table 28. It should be remembered that the correlation coefficient ranges from -1 to +1, and is a measure of the linear relationship between two variables. Coefficients greater than zero indicate that high costs are more likely to occur with high values of the other variable. The closer the coefficient is to one, the closer the linear relationship. Conversely, coefficients less than zero indicate that low costs are more likely to occur with high values of the other variable.

An extensive examination of the correlations between effectiveness and each of the various cost categories was conducted. The results of this analysis are presented by grade for each type of class in Tables 29, 30, and 31. Because of the large number of coefficients in these tables, only correlations that have an absolute value greater than 0.10 are displayed. Values smaller than this are considered to be zero because they are generally statistically insignificant.

The results of the analysis of correlation coefficients do not suggest any clear relationships between cost and effectiveness that are consistent over a variety of situations. Very few coefficients are greater than an absolute value of 0.25. At this level, less than 7 percent of the variation between effectiveness and cost can be explained by a relationship between the two.

Of additional concern is the frequency of negative correlation coefficients. Approximately half the significant relationships are negative. A negative coefficient is difficult to explain because it implies that higher levels of cost are associated with less effective programs. This might be caused by a third factor that may be influencing both the cost and the effectiveness of the programs. For example, it is possible that classes with low socio-economic students are more likely to be less effective. However, the low socio-economic nature of the classes means they are eligible for additional federal funding, and, as a result, they tend to cost more. It was shown earlier that compensatory programs funded with ESEA Title I funds are more expensive. A complete examination of the problem is not within the scope of this analysis because the socio-economic characteristics of the classes were not provided to RMC.

Table 28

CORRELATION COEFFICIENT BEWTEEN TOTAL COST AND SELECTED VARIABLES

		Compensatory Students				Noncompensatory Students		
		Conventional[b] Classrooms		Special Reading Groups		CR Schools		NCR Schools
		Only	Mixed	Only	Mixed	Only	Mixed	
Grade 2	Pre-Test	-.10	.04	.09	-.06	.17	-.06	.06
	Post-Test	.06	-.02	.05	.02	.09	-.04	-.16
	Effectiveness[a]	.24	-.07	.02	.11	-.07	-.01	-.35
Grade 4	Pre-Test	.10	-.12	-.13	.42	.03	-.12	-.03
	Post-Test	.11	-.13	-.16	.27	.04	-.11	-.05
	Effectiveness[a]	.07	-.06	-.13	-.26	.05	-.00	.12
Grade 6	Pre-Test	-.13	-.04	.09	NA	-.12	-.03	.39
	Post-Test	-.11	-.06	.10	NA	-.10	-.04	.35
	Effectiveness[a]	-.01	-.08	.03	NA	.02	-.02	.02

a. Calculated using the quadratic form, equation 3, and coefficients in Table 25.

b. As described in Chapter 2 and used in all analyses, conventional classrooms are considered as the typical classroom setting with one teacher teaching all subjects to 20 to 30 students.

Table 29

CORRELATION COEFFICIENTS BETWEEN EFFECTIVENESS[c] AND SELECTED COST CATEGORIES FOR GRADE 2[a]

	Compensatory Students				Noncompensatory Students		
	Conventional[d] Classrooms		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
1 Classroom Instructor	.29						-.27
2 Teacher Aide			-.20		-.24		-.37
3 Reading Specialist		-.11		.22			
4 Other Special Teachers	-.12		-.28	-.26			-.19
5 Nonteaching Academic Staff	-.13		-.24				-.17
6 Nonteaching Non-Academic Staff			.14	-.13			-.11
7 Supervision and Administration	.50		.26				-.32
8 Other Staff (Variables 4-7)	.42		.11				-.27
9 All Staff (Variables 1-7)	.27			.11			-.37
10 Facilities	.26						-.20
11 Tape Recorders/Record Players	-.12		-.20	.27		-.11	-.14
12 Teaching Machines	-.10		-.10	.10			-.15
13 Computers		b		b		b	b
14 Movie/Filmstrip Projectors	-.19		-.23				.15
14 Slide/Overhead/Opaque Projectors		-.16		.31			-.17
16 Television/Video Tape Recorders		-.14	-.23	-.10	-.14		
17 Other Equipment				b			b
18 All Equipment (Variables 11-17)	-.14		-.24	.23			-.10
19 Basal Readers		.13			.14		-.13
20 Individual Reading Kits			-.15		-.34		-.23
21 Library Books							-.11
22 Other Supplies and Materials	.31			.18		-.13	-.14
23 All Materials (Variables 19-22)							-.17
24 District Centers	-.26			.21			.13
25 All Resources (Variables 9, 10 18, 23, and 24)	.24			.11			-.35

a. Correlations with an absolute value less than 0.10 are not presented.
b. No data were available for comparison.
c. Calculated using the quadratic form, equation 3, and coefficients in Table 25.
d. See Note b, Table 28.

Table 30

CORRELATION COEFFICIENTS BETWEEN EFFECTIVENESS[c] AND SELECTED COST CATEGORIES FOR GRADE 4[a]

	Compensatory Students				Noncompensatory Students		
	Conventional[d] Classrooms		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed	Only	Mixed	
1 Classroom Instructor	.15		.13	-.44	.10		.25
2 Teacher Aide			-.28				
3 Reading Specialist			-.18	-.17			
4 Other Special Teachers	-.11	-.11		.19	-.13		-.14
5 Nonteaching Academic Staff			.14	.11		.10	-.10
6 Nonteaching Non-Academic Staff	.11		.29	.34		.15	
7 Supervision and Administration			.35	.25	-.13		-.14
8 Other Staff (Variables 4-7)			.18	.27	-.14		-.16
9 All Staff (Variables 1-7)			-.15	-.22			
10 Facilities	.14		-.10	-.40	.21		.25
11 Tape Recorders/Record Players			-.11	.23	-.15		
12 Teaching Machines			.21				
13 Computers		-.15		-.33	b		b
14 Movie/Filmstrip Projectors			-.18	.13			
14 Slide/Overhead/Opaque Projectors	-.12			.13			.19
16 Television/Video Tape Recorders			-.15	.14			.30
17 Other Equipment			.16	b			b
18 All Equipment (Variables 11-17)			-.10	.11	-.11		.13
19 Basal Readers					.15		.11
20 Individual Reading Kits	.19		-.29				-.21
21 Library Books		-.14	-.19		-.11		
22 Other Supplies and Materials				-.16	.16		.10
23 All Materials (Variables 19-22)	.10		-.21				
24 District Centers	-.10		.35	-.12	-.12	-.14	.30
25 All Resources (Variables 9, 10 18, 23, and 24)			-.13	-.26			.12

a. Correlations with an absolute value less than 0.10 are not presented.
b. No data were available for comparison.
c. Quadratic form. See Note c, Table 29.
d. See Note b, Table 28.

Table 31

CORRELATION COEFFICIENTS BETWEEN EFFECTIVENESS[c] AND SELECTED COST CATEGORIES FOR GRADE 6[a]

	Compensatory Students				Noncompensatory Students		
	Conventional[d] Classrooms		Special Reading Groups		CR Schools		NCR Schools
	Only	Mixed	Only	Mixed[b]	Only	Mixed	
1 Classroom Instructor			.24				
2 Teacher Aide		-.15	-.15				.20
3 Reading Specialist							-.32
4 Other Special Teachers						.14	.19
5 Nonteaching Academic Staff							.10
6 Nonteaching Non-Academic Staff			.54				
7 Supervision and Administration			.36		-.12		
8 Other Staff (Variables 4-7)			.24				.15
9 All Staff (Variables 1-7)							
10 Facilities							-.18
11 Tape Recorders/Record Players							.19
12 Teaching Machines	-.20		.21				
13 Computers	-.25		-.13				b
14 Movie/Filmstrip Projectors			-.35				.19
14 Slide/Overhead/Opaque Projectors			.31				.16
16 Television/Video Tape Recorders		.17			-.15		
17 Other Equipment		.14	b			.14	b
18 All Equipment (Variables 11-17)							.25
19 Basal Readers							.17
20 Individual Reading Kits		.13			.12		
21 Library Books							
22 Other Supplies and Materials		.19			.12		-.22
23 All Materials (Variables 19-22)		.11					.13
24 District Centers	-.24		.24		-.12		.15
25 All Resources (Variables 9, 10 18, 23, and 24)							

a. Correlations with an absolute value less than 0.10 are not presented.
b. No data were available for comparison.
c. Quadratic form. See Note c, Table 29.
d. See Note b, Table 28.

REGRESSION EQUATIONS

In an attempt to identify relationships between the various categories of cost and effectiveness, multiple linear regression techniques were used. As discussed in Chapter 5, the following functional form was employed:

$$PE = C + \Sigma a_i X_i + \varepsilon,$$

where

PE = program effectiveness,

X_i = program cost variables for each particular resource type,

C, a_i = regression coefficients, and

ε = error term.

Regression techniques allow this equation to be solved for its regression coefficients to minimize the square of the error term using the observed data.

Only the 15 class types involving various combinations of CR and NCR students in conventional classrooms were examined. There were not enough observations involving special reading groups to adequately investigate the 25 categories of cost.

The first step in the process of identifying the relationships involved including all cost categories in a regression analysis for all the classes under examination. After the results of the original regression equations were reviewed, additional regression analyses were conducted by eliminating all variables that either failed to show statistically significant relationships with effectiveness or were highly intercorrelated with each other. Further analyses also used several combinations of clusters resulting from the typal analysis. These additional analyses failed to provide consistently improved results (larger coefficients of determination) over the use of the finest level of cost variables.

After several iterations of regression analysis, a set of equations was produced containing only the resource cost variables that were significantly related to the effectiveness measure. Table 32 presents a summary of the results obtained. Only 11 equations are presented. No statistically significant equation was found that related cost and effectiveness in the following types of classrooms:

(1) second-grade conventional classrooms with NCR students only,

(2) NCR students in second-grade conventional classrooms with mixed CR/NCR students,

Table 32

STATISTICALLY SIGNIFICANT REGRESSION EQUATIONS RELATING EFFECTIVENESS[c] TO VARIOUS COST CATEGORIES[a,b]

	Grade 2					Grade 4					Grade 6				
			Mixed					Mixed					Mixed		
	CR	NCR	CR	NCR	NCR Sch	CR	NCR	CR	NCR	NCR Sch	CR	NCR	CR	NCR	NCR Sch
Class Instructor	+														
Teacher Aides	-				-									+	
Other Special Teachers									+						
Nonteaching Non-Academic Staff															
Supervision and Administration	+						+								
Facilities											-				
Teaching Machines															
Slide/Overhead/Opaque Projectors			-							+					
Television/Video Tape Recorders									+						
Basal Readers			+			+									
Individual Reading Kits								-							
Library Books													+		
Other Supplies and Materials									-	+	-				
District Centers	-														
Coefficient of Determination (R^2)	.34	N/A	.03	N/A	.12	.03	.03	.02	.05	.16	.09	N/A	.03	.02	N/A

a. The regression equation is significant at the 0.05 level. Only the signs of regression coefficients are presented. A "+" indicates a positive relationship between cost and effectiveness, while a "-" indicates a negative relationship.

b. Only conventional classrooms were examined in this analysis. See Note b, Table 28, for a definition of conventional classroom.

c. Quadratic form. See Note c, Table 29.

(3) sixth-grade conventional classrooms with NCR students only, and
(4) sixth-grade conventional classrooms in NCR schools.

The regression coefficients represent the marginal (or incremental) increase in program effectiveness that can result if one additional dollar is spent in providing that resource. This concept is based on economic theory. However, as pointed out in Chapter 5, this interpretation may not be appropriate for this analysis. Therefore, only the signs of the significant coefficients are presented in Table 32. A plus sign indicates a positive relation between the cost category and effectiveness (i.e., high program cost is associated with high program effectiveness). Conversely, a minus sign indicates a negative relation between cost categories and effectiveness (i.e., high program cost is associated with low program effectiveness). The complete equations, with the actual value of the coefficients and the appropriate statistical measures, can be found in Appendix B.

Observation shows that consistency in significant variables across strata is almost non-existent. Seldom was any particular resource cost category found to have a statistically significant relationship with effectiveness in more than one of the 11 strata for which equations were developed. However, in three cases--

- cost of teacher aides,
- cost of basal readers, and
- cost of district centers--

the costs were significant in more than one strata. The results were somewhat consistent with respect to the sign of the regression coefficient--negative for teacher aides, positive for basal readers, and usually negative (three of four) for district centers.

Table 32 also reveals that the regression equations explain very small amounts of the total variation in the effectiveness measure. This can be seen by examining the coefficient of determination. The interpretation of this coefficient, which is the square of the correlation coefficient, was explained in the analysis of variance discussion in Chapter 2. On the average, only about 9 percent of the variation in program effectiveness can be associated with the program cost. Only second-grade conventional classrooms with CR students seem to exhibit any real meaningful relationship between effectiveness and cost, with a coefficient of 0.34. Second- and fourth-grade classes

in NCR schools exhibit additional evidence of such a relationship. However, the coefficients of these equations are low at 0.12 and 0.16, respectively. The other equations have even smaller coefficients.

It is necessary to point out here that the cost of reading specialists in conventional classrooms was restricted to those cases where the reading specialist was not separately identified in connection with a class. Thus, the cost was spread over many groups of students. If the reading specialist was identified as the primary classroom instructor, she was omitted from this particular analysis.

Before we go on, we must address a major problem associated with the multiple regression methodology as it applies to the analyses conducted and presented above. In general, the correlations between our categories of resource cost were quite high, especially when compared with the correlation between any particular category of resource cost and our effectiveness measure. This situation is undesirable in regression analysis since the determined coefficients are very sensitive to small changes in the value of the dependent variable. Omission or inclusion of one cost variable in the analysis may change the sign of some other cost variable and therefore give results that are difficult to interpret.

ANALYSIS OF VARIANCE OF EFFECTIVE PROGRAMS

As discussed at length in other parts of this volume, there is some concern about the meaningfulness of the measure of effectiveness. Because some of the statistical ground rules required by the regression analysis may have been violated, the confidence that can be placed in the results has been reduced. Of serious concern is the problem of heteroscedasticity, or a systematic change in the variance of the measure of effectiveness. As observed in Figure 8 (Chapter 4), it would appear that the variance of effectiveness seems to decrease inversely with increases in the pre-test score. The maximum range of effectiveness is approximately ±40 for low pre-test scores, while the range is less than ±10 for high pre-test scores. This variation all but destroys the interval scale of effectiveness for regression analysis.

The approach that was used identified effective programs as those classes that scored substantially higher on the post-test than would normally be expected for a given pre-test. Likewise, ineffective programs were identified as those classes that scored substantially lower than would be expected.

The balance of the classes were grouped together and identified as average. Analysis of variance techniques were used to determine significant differences between the costs of the three groups. Unlike other analyses presented in this chapter, the special reading groups were not separated from the conventional classes. Thus, only five types of classes were examined. In addition, all of the cost categories were not used in the ANOVA. For reasons of simplicity, we examined only 10 of the original 25 cost categories.

To accomplish this analysis, the data for each of the five types of classes and the three grades were stratified by pre-test score. A total of 14 strata were used, each covering a range of 10 units from a pre-test score of 30 to 130. For each level of pre-test score, the post-test scores were divided into three strata corresponding to effective programs, average programs, and ineffective programs. The following general rules were used to make this division: (1) approximately 20 percent of the classes with the highest post-test scores were defined to be the most effective programs, (2) approximately 20 percent of the classes with the lowest post-test scores were defined to be the most ineffective programs, and (3) the balance of the classes were considered to be average. These ranges, while without justification, appear to be a reasonable approximation of what would be called effective and ineffective programs.

As discussed in Chapter 2, two measures result from an analysis of variance: the F test, and the R^2 statistic. A significant level of 0.05 will be used for the F test to determine when a relationship exists between cost and effectiveness.

The first series of ANOVAs was conducted for two types of students--CR students in classes by themselves and CR students in mixed classes--using the three lowest pre-test strata. Thus, attention was given to only low scoring classes with CR students. The results of this analysis are shown in Table 33. Only the situations where the results are significant at the 0.05 level are presented. If the results are significant, with the more effective programs costing more, a plus sign is used. If the results are significant, but with the ineffective programs costing more, a minus is used. The R^2s for these tests are generally fair. All but one (reading specialist in the second-grade CR only with 0.08) exceeded 0.1, with an average of 0.3. The highest was other staff in the sixth-grade CR mixed with 0.58.

Table 33

RESULTS OF THE ANALYSIS OF VARIANCE OF EFFECTIVE PROGRAMS[c]
USING ONLY CLASSES WITH LOW PRE-TEST SCORES

	Grade 2		Grade 4		Grade 6	
	CR Only	CR Mixed	CR Only	CR Mixed	CR Only	CR Mixed
Class Instructor	−[a] (.10)					
Teacher Aides			+ (.15)			
Reading Specialists	+[b] (.08)				− (.30)	− (.58)
Other Staff	− (.11)				+ (.32)	
Slide Projectors						
Tape Recorders						
Movie Projectors				+ (.52)		
Basal Readers	+ (.15)					
Individual Reading Kits						
Library		+ (.08)				

a. Significant differences exist in the cost of effective and ineffective programs at the 0.05 level, with ineffective programs costing more. The coefficient of determination (R^2) is shown in parentheses.

b. Same as Note a, except that effective programs cost more.

c. Effective programs are defined as those programs with a post-test score in the eightieth percentile or above. Ineffective programs have post-test scores at the twentieth percentile or below. See text.

The second series of ANOVAs was conducted for all five groups of students over the entire range of pre-test scores. The results of this analysis are presented in Table 34. As before, only the significant results at the 0.05 level are presented. Again, plus and minus signs are used to indicate the direction of the difference. The R^2s for these ANOVAs are much lower than observed for the previous analysis. Here, only six situations had R^2s in excess of 0.1. These were the cost of class instructors and other staff in the second-grade CR only, library books in the second-grade CR mixed, individual reading kits in second-grade NCR only, individual reading kits in second-grade NCR schools, and class instructors in fourth-grade NCR schools. The average R^2 of these six is 0.24. A more complete presentation of these results can be found in Appendix C.

INDEPENDENT ANALYSIS OF POST-TEST SCORES

This analysis is another attempt to examine the relationship between cost and effectiveness without using the measure of effectiveness calculated in Chapter 4. In the previous analysis, the differences in cost between classes that scored higher on the post-test and classes that scored lower on the post-test were examined. This analysis is similar, but with a slightly different thrust. Here the differences in post-test scores were examined for classes that spent more or less money.

To conduct the cost-effectiveness analysis, the 15 conventional classroom data sets were divided further into groups representing high, medium, and low total resource costs. Each of the three groups had an approximately equal number of observations. Separate regression analyses were then conducted using the quadratic form presented in equation 3 (Chapter 4).

In addition, the groups of students taught by reading specialists were subjected to separate regression analyses. The resulting curves were plotted as dashed lines in the figures for which comparable curves had previously been plotted. However, the means of the total cost per student in these special reading groups were considerably higher (usually by a factor of two) than the mean cost for the data sets defined as high cost conventional classrooms.

Table 34

RESULTS OF THE ANALYSIS OF VARIANCE OF EFFECTIVE PROGRAMS[c] USING ALL CLASSES

	Grade 2					Grade 4					Grade 6				
			Mixed					Mixed			Mixed				
	CR	NCR	CR	NCR	NCR Sch	CR	NCR	CR	NCR	NCR Sch	CR	NCR	CR	NCR	Sch NCR
Class Instructor	-[a] (.13)					- (.05)				- (.30)					
Teacher Aides			- (.07)						- (.06)						
Reading Specialists			- (.08)	+ (.01)		- (.11)			- (.07)					- (.07)	
Other Staff	+[b] (.21)			+ (.03)							+ (.06)			+ (.30)	
Slide Projectors		+ (.04)		- (.04)		+ (.03)						+ (.09)		+ (.13)	- (.33)
Tape Recorders			- (.03)	- (.04)					- (.01)						
Movie Projectors					- (.25)										
Basal Readers				+ (.03)		- (.09)								+ (.06)	
Individual Reading Kits		- (.14)		- (.05)					- (.12)						
Library Books			- (.20)	- (.04)	+ (.16)	- (.04)						- (.09)		- (.06)	

a. Significant differences exist in the cost of effective and ineffective programs at the 0.05 level, with ineffective programs costing more. The coefficient of determination (R^2) is shown in parentheses.

b. Same as Note a, except that effective programs cost more.

c. Effective programs are defined as those programs with a post-test score in the eightieth percentile or above. Ineffective programs have post-test scores at the twentieth percentile or below. See text.

The resulting regression curves, presented in Figures 10 through 24, were not plotted beyond the range for which mean pre-test scores were available for each specific regression analysis. Separate figures were used for each primary data set, with each figure representing relatively homogeneous treatment types except for cost level. These figures allow visual evaluation of the mean post-test scores for various levels of both pre-test scores and total resource cost.

This method does not require the test scores to be standardized, nor for the test to be nationally normed. The only requirement is that pre- and post-test scores be measured by "equivalent" tests for all groups being compared. Figures 10, 11, 12, 16, and 17 represent primary strata that contain the five secondary strata made up of units taught largely by reading specialists. The plotted curves for these five strata are represented by dotted lines. All solid lines represent strata made up of classes taught by regular teachers.

A general analysis of all 15 figures with respect to the solid regression lines shows that higher costs cannot be strongly related to higher achievement (as measured by the test used for this study). In fact, in as many instances as not, substantial portions of the regression lines representing achievement in high cost classrooms fall below regression lines representing achievement in lower (medium and low) cost classrooms.

It is important here to point out that emphasis must not be placed on any particular regression curve. The sample of classes on which these regressions are based vary considerably in size (from 12 to 106).

The calculations required to identify confidence limits around all 50 of the regression curves are quite complex and would be very time-consuming. However, it was found that the standard errors were generally between two and five units from the curves at the pre-test means. Obviously, at the outer limits of the curves, the error is more. This range of standard error implies, at least in some instances, that differences at specific points on two specific curves would be approaching statistical significance. However, significant differences over large portions of any two specific curves are rare. In addition, the results are certainly not consistent across all strata.

The following example demonstrates why specific curves should not be used as a basis for decisions. In Figure 10, the differences between the high cost curve (H) and the other curves at the low end of the pre-test scale

are quite different (at least 20 test points). However, during the later stages of the analysis, it was discovered that 18 of the data points on which the upper end (indicating high achievement) of the curve was based were developed using data from the same school. Investigation showed that the school was experimental in nature and had a large number of reading specialists who were not identified in the data collection efforts.

Observation of the regression curves where special reading groups were separated from the conventional classes shows that the post-test scores for the special reading groups are substantially above the other curves. This is true for all five cases, which implies rather conclusively that special reading groups demonstrate higher achievement (as measured by the tests) than classes in other strata. In many instances, the difference between the curves is in excess of the standard error range.

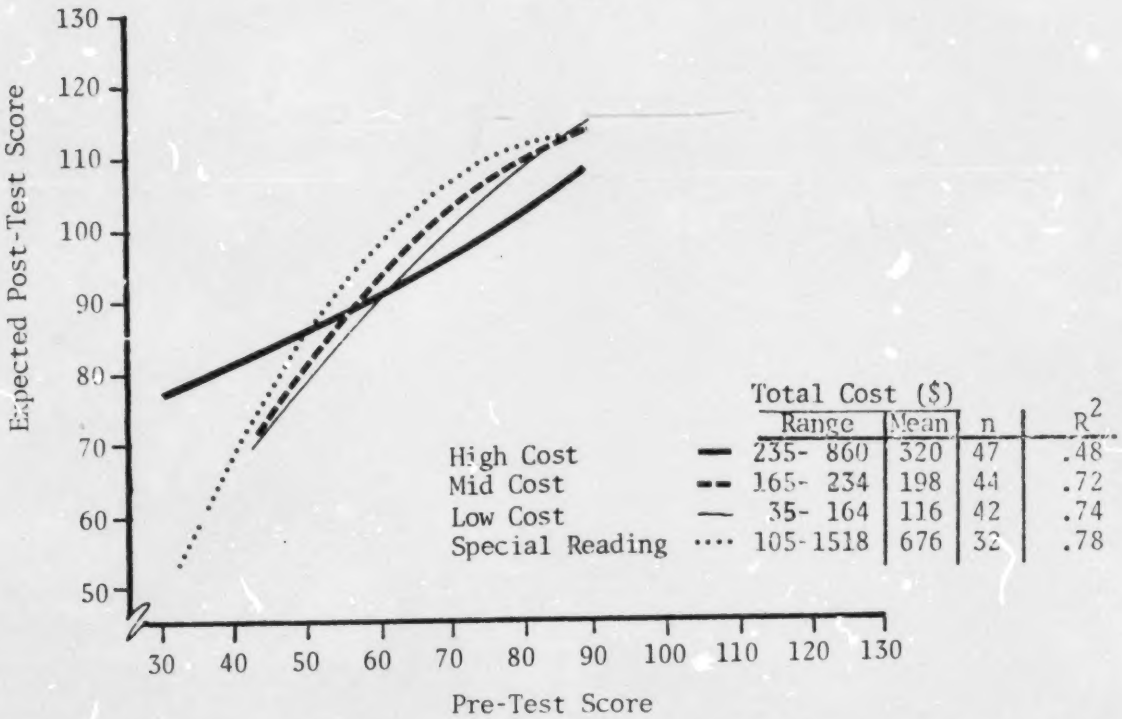


Figure 10: SECOND-GRADE COMPENSATORY GROUPS, EXCLUDING GROUP SIZES LESS THAN FOUR

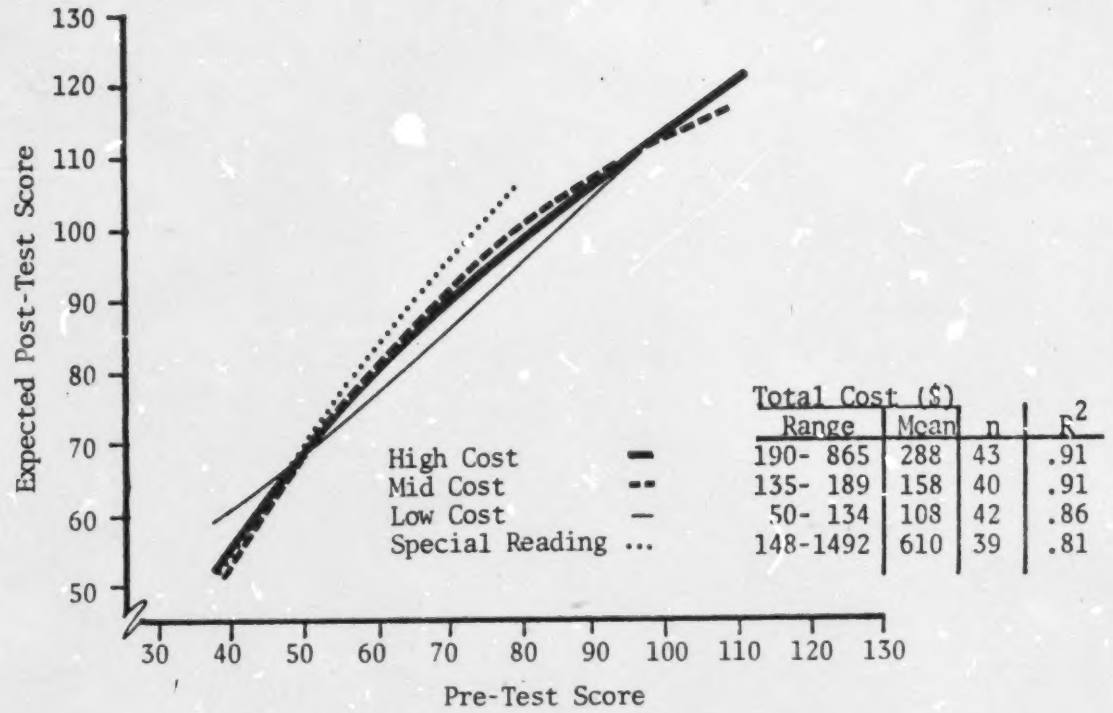


Figure 11: FOURTH-GRADE COMPENSATORY GROUPS, EXCLUDING GROUP SIZES LESS THAN FOUR

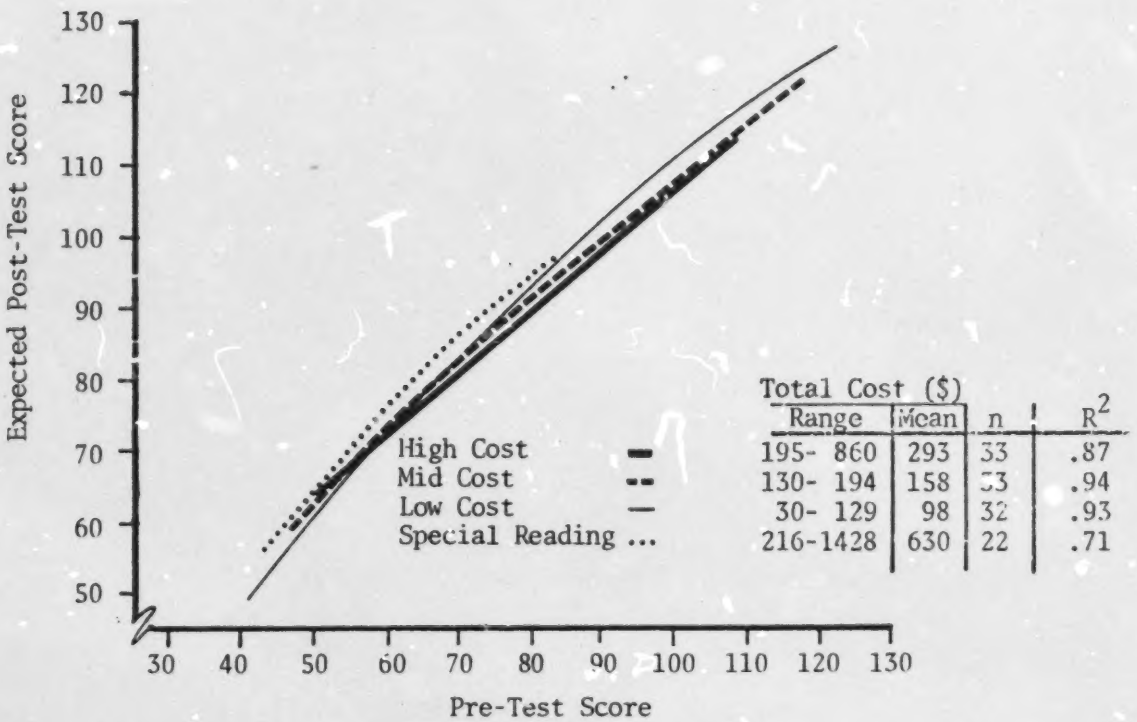


Figure 12: SIXTH-GRADE COMPENSATORY GROUPS, EXCLUDING GROUP SIZES LESS THAN FOUR

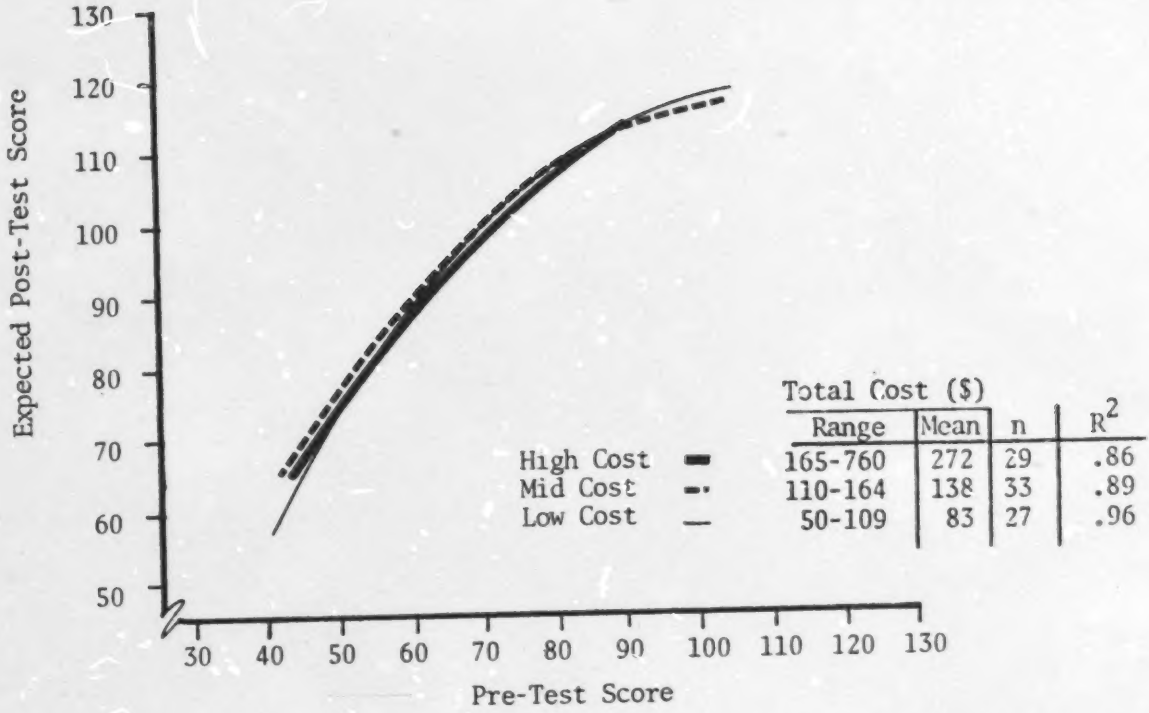


Figure 13: SECOND-GRADE NONCOMPENSATORY GROUPS IN SCHOOLS WITH COMPENSATORY CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

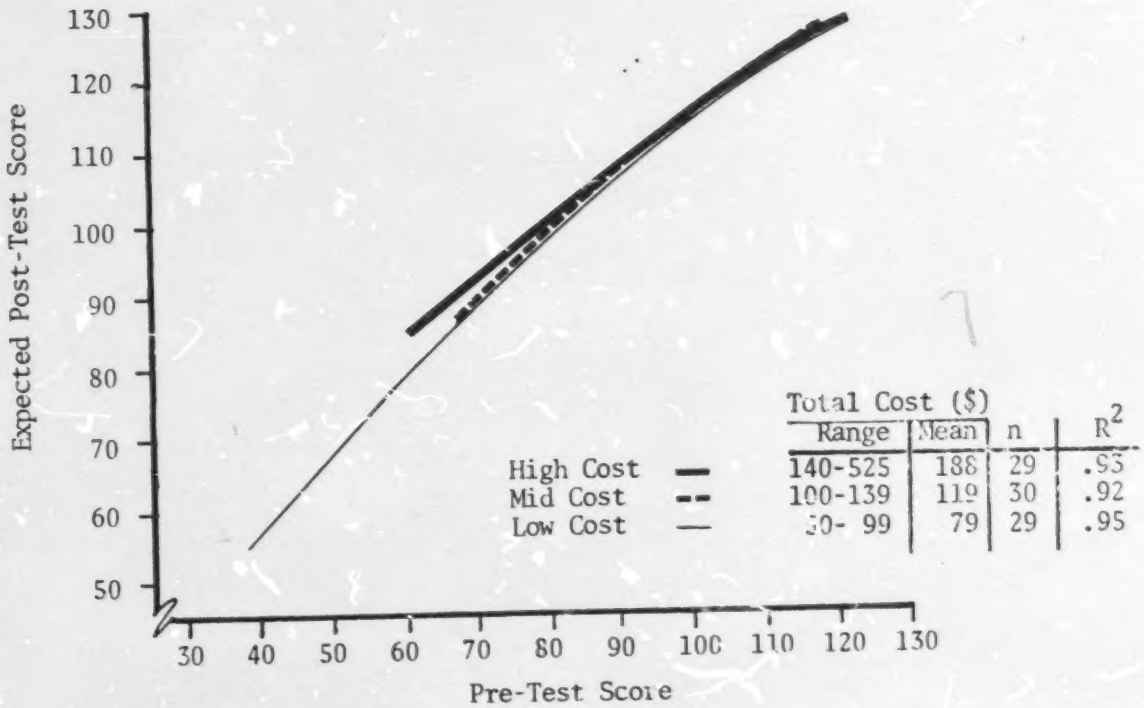
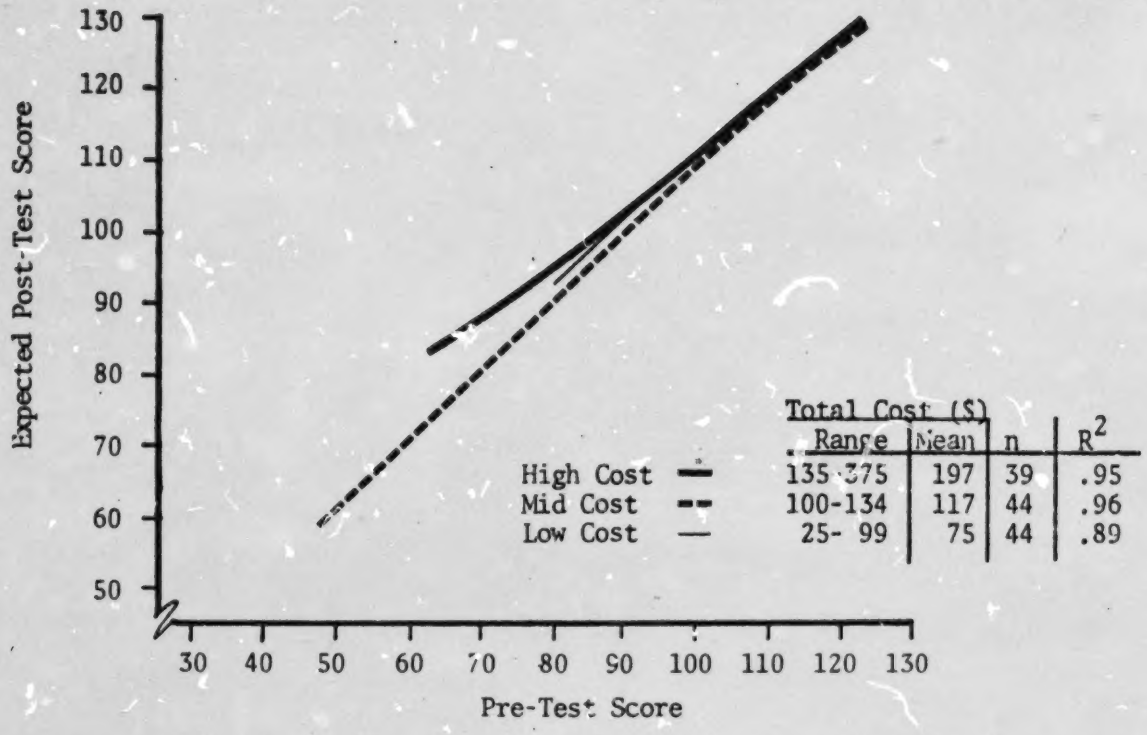
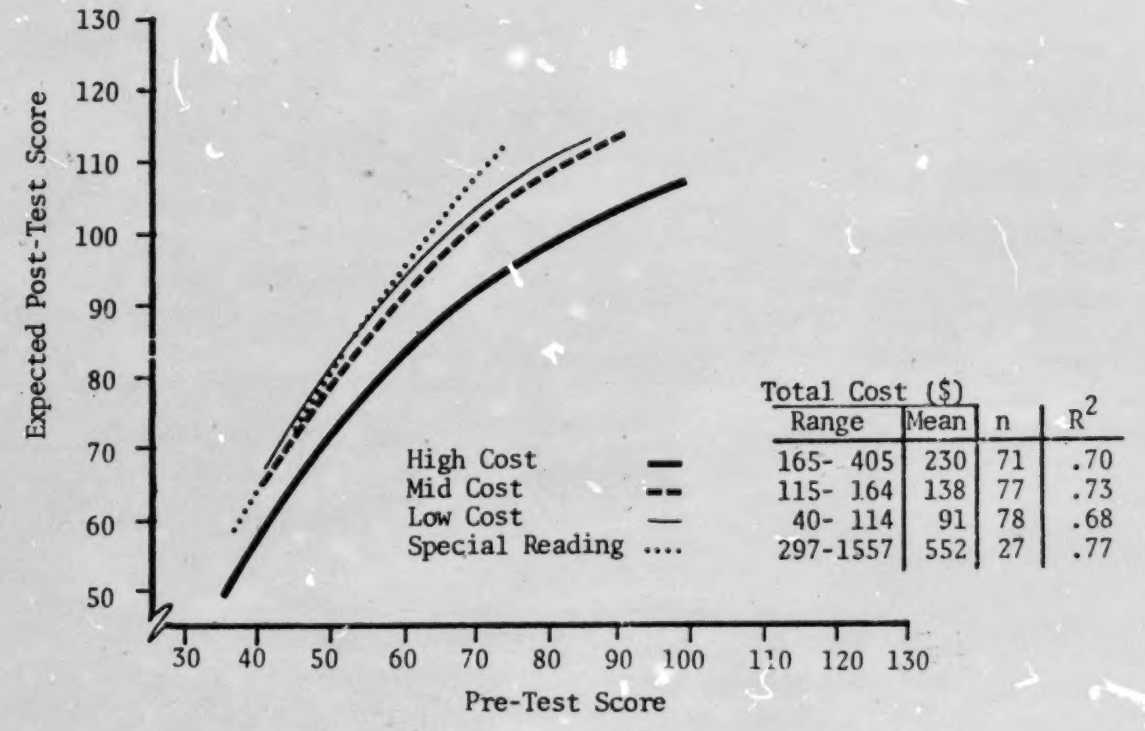


Figure 14: FOURTH-GRADE NONCOMPENSATORY GROUPS IN SCHOOLS WITH COMPENSATORY CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS



	Total Cost ($) Range	Mean	n	R^2
High Cost	135-375	197	39	.95
Mid Cost	100-134	117	44	.96
Low Cost	25- 99	75	44	.89

Figure 15: SIXTH-GRADE NONCOMPENSATORY GROUPS IN SCHOOL WITH COMPENSATORY CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS



	Total Cost ($) Range	Mean	n	R^2
High Cost	165- 405	230	71	.70
Mid Cost	115- 164	138	77	.73
Low Cost	40- 114	91	78	.68
Special Reading	297-1557	552	27	.77

Figure 16: SECOND-GRADE COMPENSATORY GROUPS IN MIXED CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR

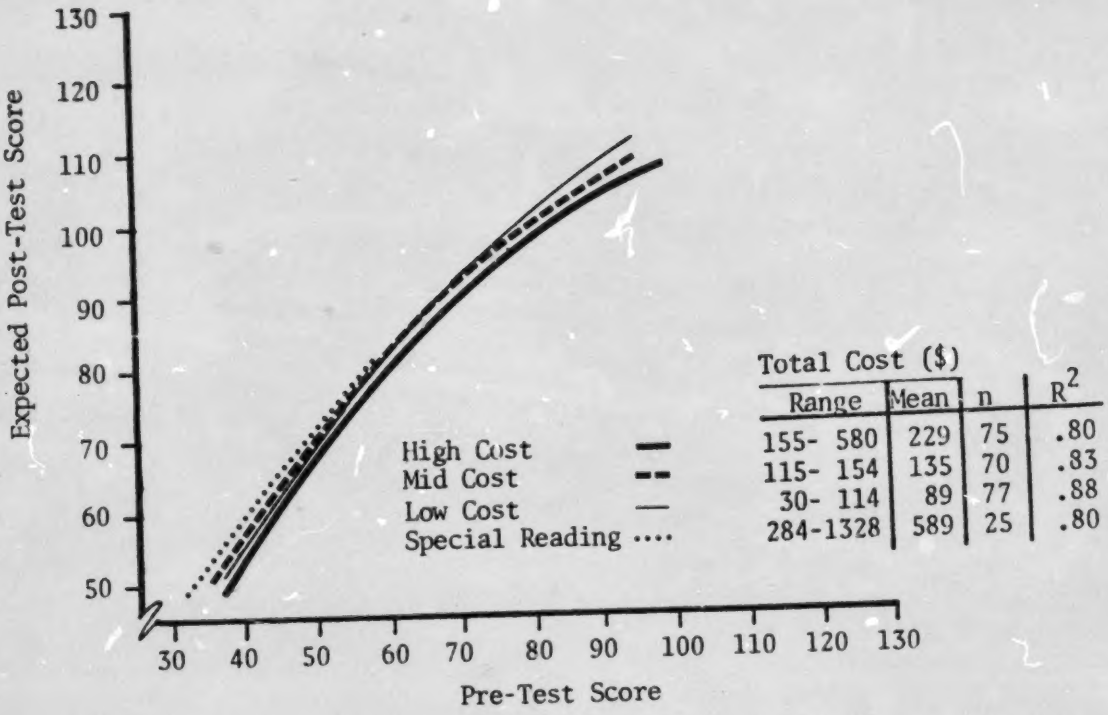


Figure 17: FOURTH-GRADE COMPENSATORY GROUPS IN MIXED CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR

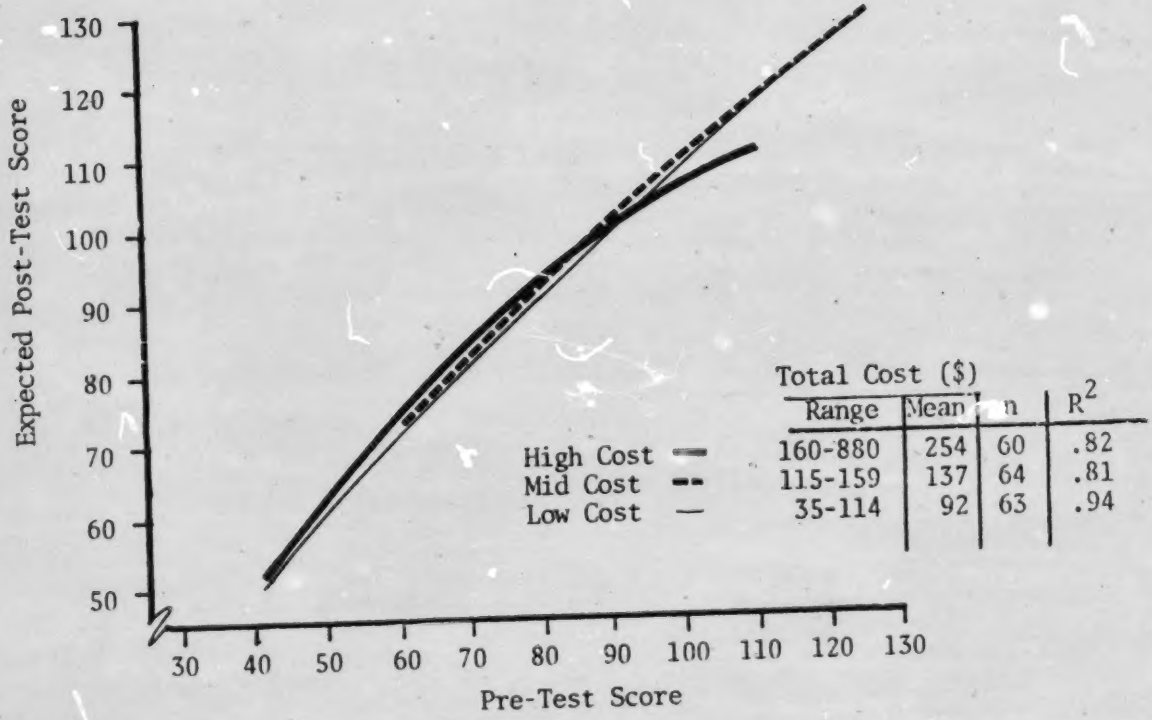


Figure 18: SIXTH-GRADE COMPENSATORY GROUPS IN MIXED CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

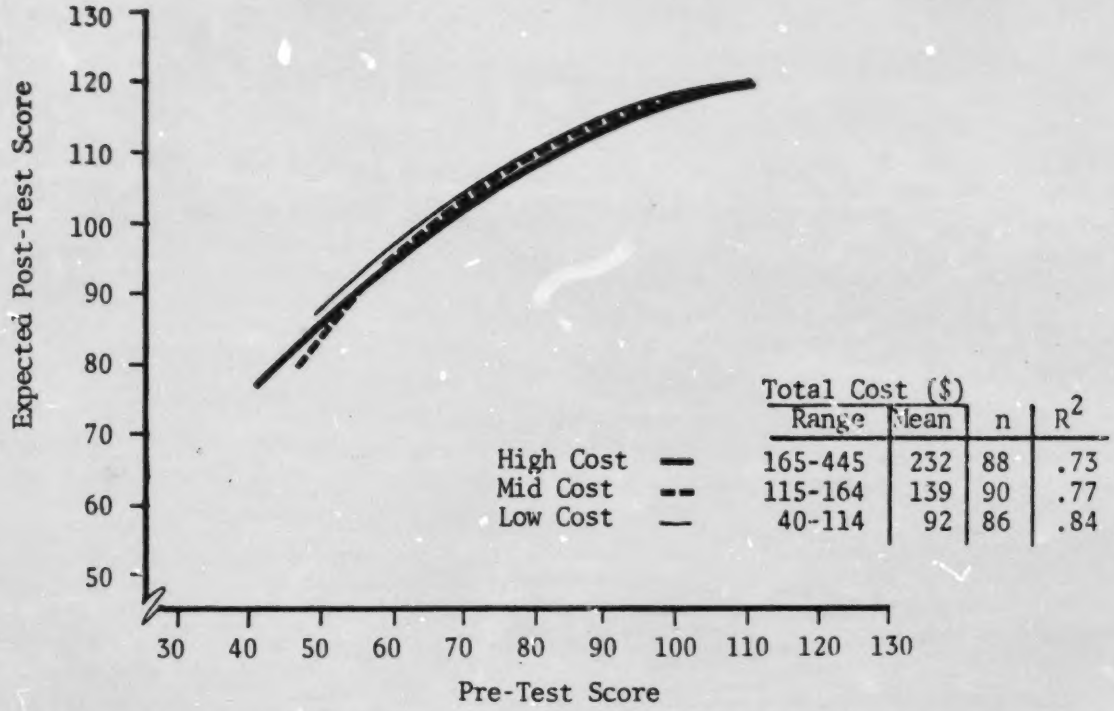


Figure: 19: SECOND-GRADE NONCOMPENSATORY GROUPS IN MIXED CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

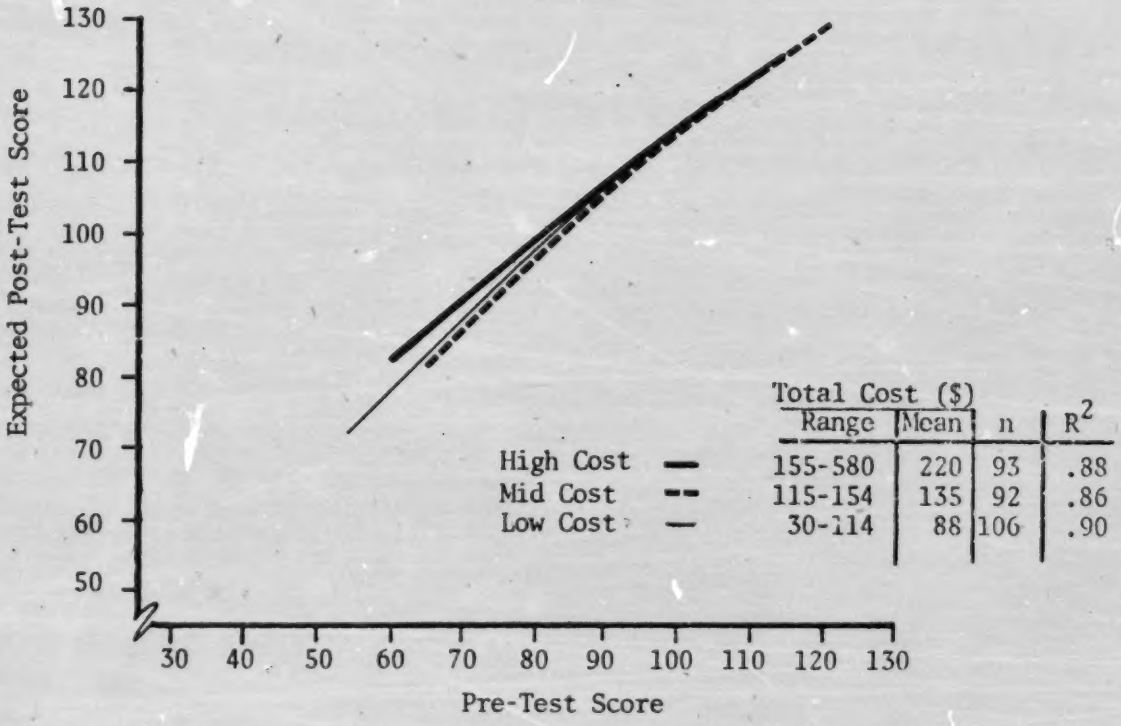


Figure 20: FOURTH-GRADE NONCOMPENSATORY GROUPS IN MIXED CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

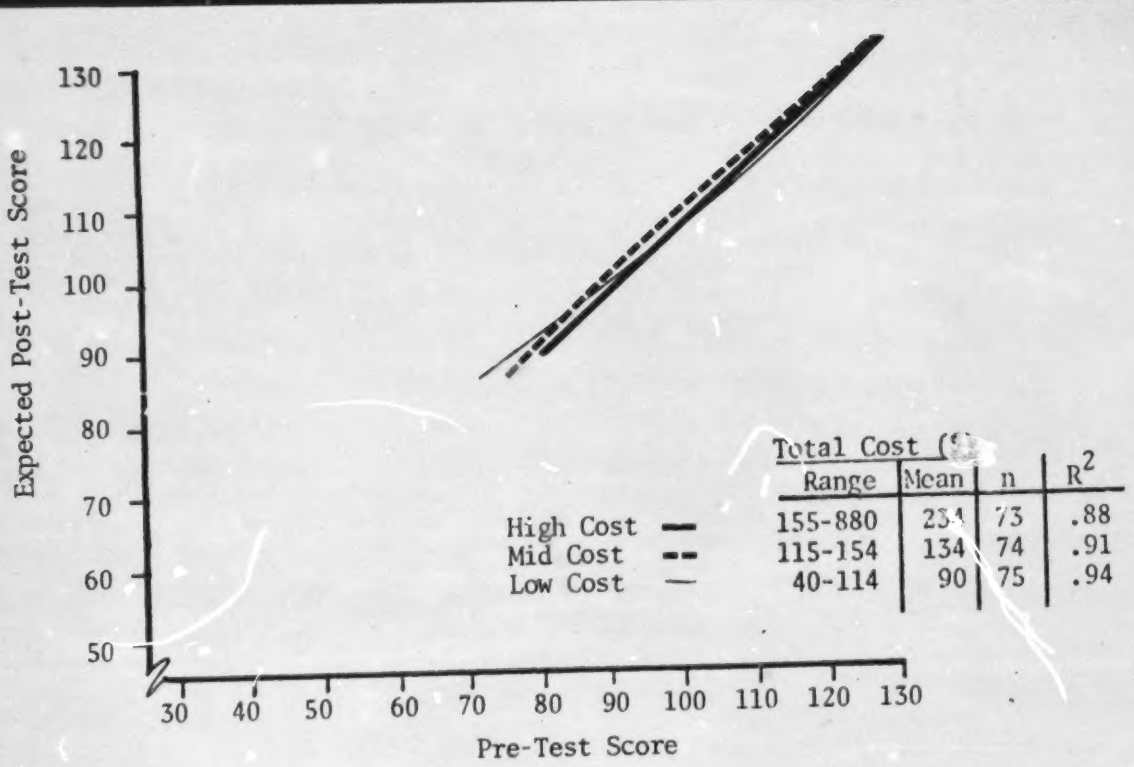


Figure 21: SIXTH-GRADE NONCOMPENSATORY GROUPS IN MIXED CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

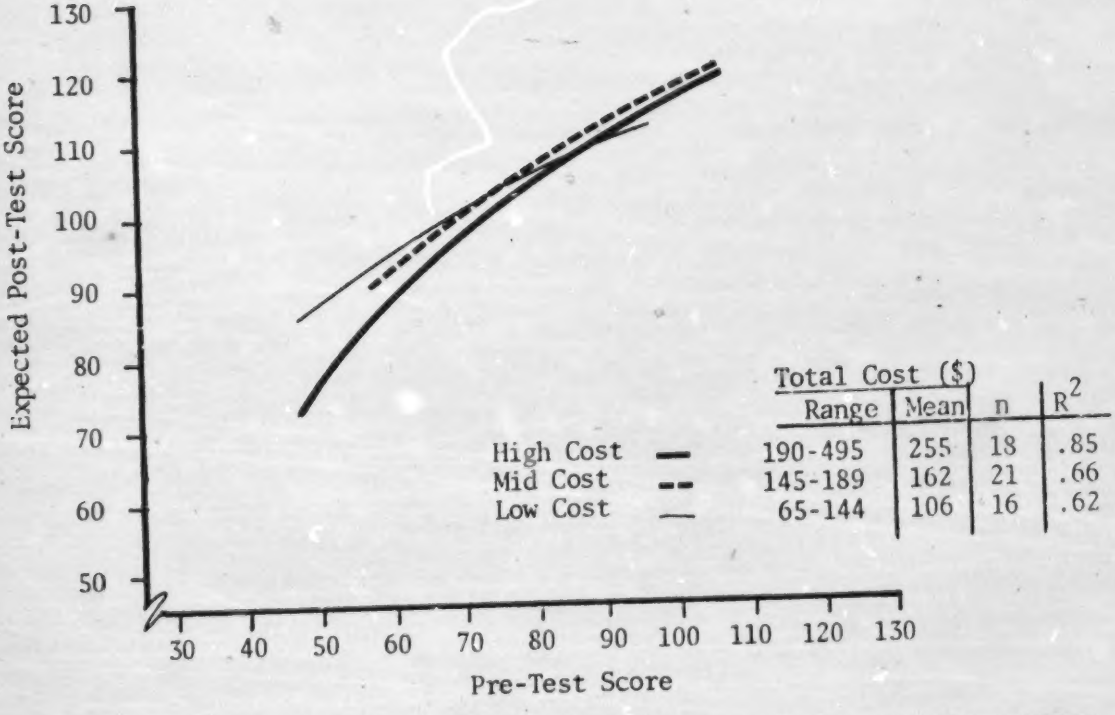


Figure 22: SECOND-GRADE GROUPS IN SCHOOLS WITH NO COMPENSATORY CLASSES, EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

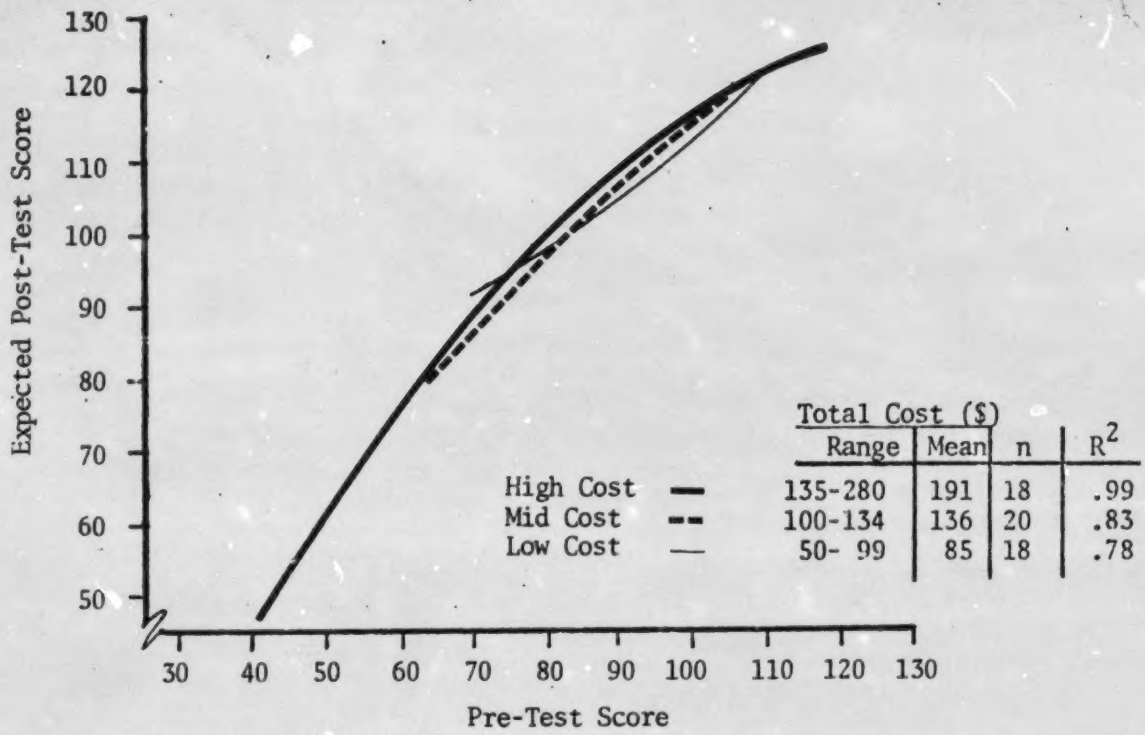


Figure 23: FOURTH-GRADE GROUPS IN SCHOOLS WITH NO COMPENSATORY CLASSES EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

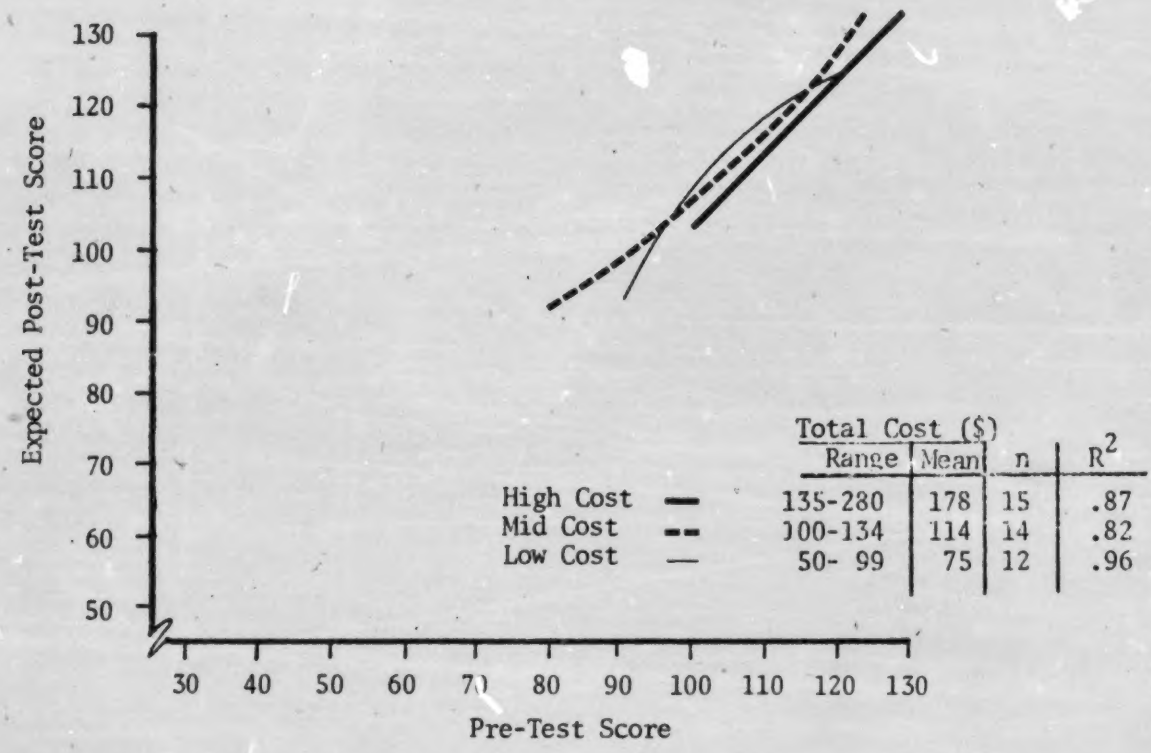


Figure 24: SIXTH-GRADE GROUPS IN SCHOOLS WITH NO COMPENSATORY CLASSES EXCLUDING GROUP SIZES LESS THAN FOUR AND GROUPS TAUGHT BY READING SPECIALISTS

SIMPLE COMPARISON

As you will recall, the analyses previously conducted in Chapters 2, 3, and 4 considered the characteristics and costs of the various class types. In an effort to uncover any additional information on the relationship between the cost and the effectiveness of the programs, these analyses were re-examined. Unfortunately, Chapter 2 determined that, in general, there were significant differences in both pre-test and post-test scores for the various types of classes.

Because of this general lack of comparability, it is impossible to conduct a simple comparison between class types. The effect of the initial conditions (as measured by pre-test) cannot be separated from the effects of the educational treatment. Thus, no conclusions can be reached.

In comparison, however, two groups of classes showed no significant differences in either the pre-test or post-test scores. Therefore, it is possible to develop a potentially interesting examination. As shown in Tables 7 and 8 (Chapter 2), these two exceptions are:

- CR students in special reading groups, either alone or mixed with NCR students, and
- CR students in conventional classrooms, either alone or mixed with NCR students.

It should be remembered that there are significant differences in both pre- and post-test scores for CR students in conventional classrooms and special reading groups with CR students. However, in either a conventional classroom or a special reading group, the presence or absence of NCR students does not appear to significantly affect the achievement of the CR students. In both cases, the pre-test scores for classes with CR students alone and those with NCR students were similar, indicating similar initial levels of achievement. Likewise, the post-test scores for classes with CR students alone and those with NCR students were similar, indicating that the effect of the educational treatment on achievement was similar.

These results differ somewhat from the ETS finding that CR students in mixed classes performed better than CR students alone. However, ETS did not separate special reading groups from conventional classrooms, as was done in this analysis. Thus, the class types are not identical and not directly comparable.

After discovering these facts, the costs for the various classes were examined using the results presented in Table 20 (Chapter 3). These results indicate that the cost of special reading groups are essentially the same whether or not NCR students are in the group. However, since the treatment effects appear similar, the fact that the cost of providing instruction to special reading groups is not affected by the presence of NCR students is expected.

The same conclusion is not true, however, for CR students in conventional classes. From Table 20 again, the cost of classes with CR students only is significantly more expensive than classes with a mix of CR and NCR students. Classes with CR students alone range from 15 percent to 34 percent more, on the average. Thus, while the treatment effects are similar (because pre-test and post-test scores are similar), there is no obvious benefit resulting from the increased cost.

The meaning of this result is difficult to interpret. It does not necessarily mean that there is no relationship between cost and program effectiveness. However, it does suggest that other factors influence program cost that have no effect on program effectiveness.

While the implications of ESEA Title I were not examined in detail, there does not appear to be strong evidence that the availability of ESEA Title I funds, and their associated policy implications, influences the cost differences between the two types of conventional classes with CR students. Examination of Tables 4, 5, and 6 (Chapter 2) shows that classes with CR students only are less likely to be fully funded from ESEA Title I funds than mixed classes. When either full or partial funding from ESEA Title I is considered, the situation is reversed. However, since the majority of both class types are receiving some ESEA Title I funds, it is unlikely that the availability of ESEA Title I funds produces the observed differences in total cost. This question will be addressed in greater detail in Addendum 2 of ETS's Volume I.

SEPARATE EXAMINATION OF UNUSUALLY EFFECTIVE PROGRAMS

An analysis of variance was conducted for the total per student reading costs associated with classes in the five schools judged by ETS to be unusually effective. This judgment was based on an examination of the results of the pre- and post-tests and on actual classroom observation. The purpose of this analysis was to determine if the costs associated with these programs were significantly different from the costs associated with other similar classes.

The results of the analysis of variance are presented in Table 35. The mean costs and their standard deviations are presented for both the unusually effective programs and the other programs. Although some differences in cost can be observed between the two groups, the differences were found to be statistically insignificant. These results indicate that the unusually effective programs are not consistently more or less costly than the other programs. Rather, on the average, they cost about the same as the others.

Table 35

ANALYSIS OF VARIANCE IN COST BETWEEN UNUSUALLY EFFECTIVE[d] PROGRAMS AND OTHER PROGRAMS

		Effective Programs		Other Programs			
		Mean Total Cost[a]	Number of Observations[b]	Mean Total Cost[a]	Number of Observations	F[c]	R^2
Grade 2	Conventional Classrooms[e] with CR Students Only	270.54 (69.72)	8	212.00 (108.33)	125	2.26	.017
	Conventional Classrooms with a mix of CR/NCR Students	139.62 (39.18)	7	151.01 (69.79)	219	.19	.001
Grade 4	Conventional Classrooms with CR Students Only	198.67 (74.70)	5	185.26 (117.49)	120	.06	.000
	Conventional Classrooms with a mix of CR/NCR Students	139.68 (37.43)	8	150.91 (75.41)	214	.18	.001
Grade 6	Conventional Classrooms with CR Students Only	132.98 (16.44)	5	186.46 (116.14)	93	1.04	.011
	Conventional Classrooms with a mix of CR/NCR Students	140.85 (31.40)	4	159.81 (97.34)	183	.15	.001
All Grades	Conventional Classrooms with CR Students Only	212.37 (82.94)	18	195.48 (114.17)	338	.385	.001
	Conventional Classrooms with a mix of CR/NCR Students	139.90 (34.98)	19	153.59 (80.73)	616	.543	.001

a. Standard deviations are displayed in parentheses below the mean.

b. The number of classrooms by type in the five schools judged to have particularly effective programs.

c. No statistically significant differences were found at the 0.05 level.

d. These programs were identified by ETS as being rated high on both the test results and the observational data. All grade levels in each school were considered in these calculations, even though ETS found grade levels in some of the schools that were not particularly effective. A separate analysis where these grade levels were excluded did not significantly affect the results.

e. As defined in Chapter 2 and used in all analyses, conventional classrooms are considered as the typical classroom setting with one teacher teaching all subjects to 20 to 30 students.

7

CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the results and conclusions developed as part of the cost-effectiveness analysis of compensatory reading programs in this volume. In addition, the methodological problems encountered in the course of this study are highlighted and discussed. Specific recommendations are also made about the direction of future studies of this nature.

SPECIFIC CONCLUSIONS

Before answering the major questions posed in Chapter 5, it seems appropriate to summarize the specific conclusions reached as a result of the following six approaches used in the cost-effectiveness analysis:

(1) an examination of correlation coefficients,

(2) the identification of relationships through linear regression techniques,

(3) an analysis of variance examination of the cost of effective programs,

(4) an independent analysis of post-test scores at various cost levels,

(5) a simple comparison of class types, and

(6) a separate examination of unusually effective programs.

The conclusions that can be drawn from each of these analyses are presented below.

Correlation Coefficients

The results of the analysis of correlation coefficients do not suggest any clear relationships between cost and program effectiveness that are consistent over a variety of situations. The coefficients are generally weak, with the vast majority under an absolute value of 0.25. A value of 0.24 indicates that less than 7 percent of the variation between cost and

effectiveness can be explained by any relationship between the two. In addition, approximately half of the coefficients have negative values that are difficult to interpret. Most of the coefficients were not significantly different from zero. Much of the lack of meaning associated with the correlation coefficients may have resulted from the weaknesses of the measure of effectiveness pointed out in the preceding chapters.

Regression Equations

An examination of 15 types of classes revealed that the functional relationships between program effectiveness and the cost of 25 categories of program resources are at best very weak. In a particular sense, almost no significance can be attached to the equations developed in this analysis. Only one of the many equations examined indicated that a significant percentage of the variability in effectiveness was related to the various costs associated with the class. This was the second-grade conventional classrooms for CR students only, which had an R^2 equal to 0.34. In only rare situations were the effects of resources consistent across the various types of classes.

Based on these results, it appears that no functional relationships exist between the resources available to a program and its effectiveness. However, considering the weaknesses of the measure of effectiveness used in this analysis, particularly the problem of heteroscedasticity, very little confidence can be placed in this conclusion. There may, in fact, be relationships between cost and effectiveness that are hidden because of our inability to meaningfully compare classes with different pre-test scores.

Analysis of Variance of Effective Programs

In an effort to overcome the problem of heteroscedasticity in the measure of effectiveness, an alternative analysis was conducted. In this analysis, the classes were divided into small groups based on pre-test scores. Each group was subdivided into effective, average, and ineffective programs. Effective programs were defined as those programs with a post-test score in the eightieth percentile or above, while the ineffective programs had a post-test score below the twentieth percentile. The costs of these groups were then examined through analysis of variance techniques. Unlike the other analyses in this report, the special reading groups were not separated from

the conventional classes when the analysis was conducted. Separate analyses were conducted for classes with low pre-test scores and all classes.

Many contradictions can be seen between the results from this analysis and those from the regression equations. These contradictions cannot be resolved because two separate approaches were used with different objectives. Despite the apparent conflicts, however, both analyses support the argument that significant relationships between cost and program effectiveness do not exist. Both analyses show very low values for the R^2 statistic, indicating that other factors influence the effectiveness of a class far more than the quantity of resources available. Therefore, it is entirely possible that spurious results are responsible for the presence of false relationships between particular cost categories and effectiveness. This, in fact, could explain the apparent contradiction between the two analyses.

Independent Analysis of Post-Test Scores

As another approach, the relationships between cost and program effectiveness were examined by dividing each type of class into groups representing high, medium, and low total costs. The expected post-test scores were calculated for each category to determine which groups had the highest average post-test score at each pre-test level. This procedure examined the relationships without using the measure of effectiveness developed in Chapter 4.

In general, no significant differences were found between the post-test scores for classes at the various cost levels. This supports the previous findings of no significant relationship between cost and program success. However, there is an exception to this generalization. Generally, higher post-test scores were found in special reading groups when compared with conventional classrooms.

Unfortunately, the cost associated with the special reading groups is quite high, generally two to three times more than the conventional reading program. Whether or not this additional cost is justified cannot be assessed in this study partly because we do not have a meaningful measure of effectiveness that can be compared at all levels of pre-test scores and partly because this is a value judgment best left to school administrators.

Simple Comparison

As a final examination, we made a systematic review of the various characteristics, cost, and effectiveness data. The major conclusion reached in

this analysis was that CR students--both alone and in mixed classes--appeared to achieve the same educational gains starting from the same pre-test levels.[1] However, the cost associated with conventional classrooms with CR students only was between 15 and 45 percent higher than conventional classrooms with a mix of CR and NCR students. Again, this supports the general conclusion that there is no relationship between cost and program success.

While this study indicates where the additional money is being spent (in this case, the additional costs are associated with increased use of teacher aides and reading specialists), it is not possible to explain why it was spent. Obviously, other factors affect cost and effectiveness that have not been fully explored in this study. Some of these factors may involve value judgments made by local school administrators based on variables not considered in this study.

While programs funded by ESEA Title I were demonstrated to cost more than other programs (see Volume I), the availability of ESEA Title I funds does not appear to be one of the factors that influence the higher costs associated with conventional classrooms having CR students only. Tables 4, 5, and 6 in Chapter 2 indicate that fewer conventional classrooms with CR students only are fully funded by ESEA Title I than conventional classrooms with a mix of students.

A Separate Examination of Unusually Effective Programs

In-depth observations were conducted by ETS at 29 schools in the sample to determine the specific characteristics of their programs. The trained observers were not informed of the results of the pre- and post-tests. Five schools were identified as being particularly effective. This selection was based on the observational data and the test results. An analysis of the total costs of the classes in these five schools indicated no significant differences between them and the other schools included in the analysis. The fact that the costs of effective programs are not significantly different from the costs of average programs lends support to the general conclusion that effectiveness is not functionally related to the resources available to a compensatory reading program (and thus its costs).

---

1. These results differ from the ETS finding that CR students in mixed classes did better than CR students alone. However, the basis of analysis used for this examination was different, and the results are not directly comparable. See Chapter 6 for a discussion of the difference.

GENERAL CONCLUSIONS

As discussed in Chapter 5, the cost-effectiveness analysis was to answer two major questions:

- If functional relationships exist between the cost of program resources and student achievement gains, what is the nature of these relationships?
- Are certain compensatory reading programs more cost effective (i.e., a greater return in achievement gain for each dollar spent) than others?

Based on the measure of effectiveness developed by ETS for use in this study, no significant relationships were found between cost and effectiveness that were consistent for a variety of similar classes. This examination included an investigation of all cost categories. However, this result may have been produced by a general weakness in the measure of effectiveness rather than a true lack of relationship. Nevertheless, the general conclusion was supported by the additional analyses conducted without the formal measure of effectiveness. In these analyses, no consistent relationship was observed between the total cost of a program and its apparent effectiveness. Thus, in response to the first question, we must offer a qualified no to the existence of a clear functional relationship between the cost of resources for compensatory reading programs and student achievement.

The response to the second question is even more difficult. No real distinction was made between various compensatory reading programs. This, in fact, may be one of the largest limitations of our study. It is entirely possible that the success of a program is a function of how resources are used rather than how much of the resources are available. For example, let's consider teacher aides. In some classes, they are little more than clerks, while in other classes they serve as important extensions of the teacher's abilities. This study made no attempt to differentiate between their use in the classroom.

Nevertheless, if we consider only the types of classes as the program descriptors, certain statements can be made regarding the various compensatory reading programs. With the possible exception of special reading groups, it has been shown that no class type is consistently more effective than another. In fact, the treatment received by CR students in conventional classes by themselves is essentially identical to that received by CR

students in conventional classes where CR and NCR students are mixed. However, the classes with CR students only cost between 15 and 45 percent more, implying that the most cost-effective programs are those that cost the least. Thus, there is very little justification for concluding that the more resources a program has, the more effective it is.

This is not to say that costs can be completely ignored, since instruction cannot take place if the resources are below some critical point. Unfortunately, it is not possible to identify this critical point with the data collected for this study. There did not appear to be any classes that were both ineffective and "resource poor." However, this may be more the result of the demonstrated relationship between low achievement and the availability of resources caused by the impact of ESEA Title I funds. Thus, no firm statement can be made about a critical level of resources.

The only possible exception to the general rule that program effectiveness seems to be influenced by the availability of resources is the special reading group. Both the analysis of effective programs and the independent analysis of post-test scores indicate that significant gains are achieved by special reading groups. In fact, these gains are in excess of those achieved by more conventional compensatory programs. However, the special reading groups are quite expensive, generally two or three times that of the conventional classes. Whether or not this additional cost is justified cannot be assessed because of the lack of a meaningful measure of effectiveness and the inherent danger of making value judgments with incomplete data.

In summary, no truly definitive conclusions can be reached regarding the cost effectiveness of alternative compensatory reading programs. The general weakness of the measure of effectiveness--that is, its inability to allow meaningful comparisons between classes with different pre-test scores--limits the confidence that can be placed in any conclusions. However, there are indications that no specific types of resources, or combinations of resources, significantly influence program effectiveness. Instruction provided in special reading groups in addition to the regular reading program appears to be the only exception to this general statement. The costs of such gains, however, are high.

The question of why there is no consistent relationship between the cost of a program and its apparent effectiveness was not really addressed in this study. However, we can conjecture possible reasons. First, and

most obvious, the availability of a resource may be less important than how it is used. For example, teacher aides vary from little more than clerks to important extensions of the teacher's ability. More importantly, there may be factors that affect student achievement far more significantly than the availability of resources, as measured in terms of dollars and cents. Factors such as parent interest, teacher and staff attitudes, or school environment may be far more important. While it may be true that the availability of resources may be indicative of these factors, or actually influences them, there do not appear to be any consistent relationships indicating that greater levels of resources improve a program.

METHODOLOGICAL PROBLEMS

Several methodological problems were encountered in this study. While there are indications that no significant relationship exists between program cost and program effectiveness, this conclusion may be a direct result of these problems. Therefore, each problem is discussed below.

Measure of Effectiveness

A residual gain score was used as the measure of effectiveness in this study. This measure is the difference between the observed post-test score and the post-test score that each class might expect to achieve based on its pre-test score. In theory, the residual gain score should have proved to be a meaningful measure of effectiveness. However, the measure of effectiveness had several inherent weaknesses that significantly reduced its usefulness.

The first was the use of test scores for all students who were ever enrolled in the class, even if both pre-test and post-test scores were not available. The number of situations where only one score was available does not represent a large portion of the total observations. Nevertheless, such situations introduce effects that are neither the result of the educational approach nor a function of the initial class achievement level. Although these effects are not readily apparent in the data, their presence tends to reduce the overall confidence in the results.

The second weakness of the measure of effectiveness lies in its general inability to provide a basis for meaningful comparisons between classes with different pre-test scores. In particular, the interval scale used to interpret the measure of effectiveness varies with the pre-test score. It was

shown that the sensitivity of the measure decreased by a factor of five between pre-test scores of 50 and 100 for the second-grade test. This weakness created analytical problems that made it nearly impossible to detect relationships between cost and effectiveness.

Program Characteristics

No major effort was made in the cost-effectiveness study to include descriptions of the educational approaches actually being employed in the classes. Since these program characteristics relate directly to how resources are used in the class, the lack of such information could hide any relationships between cost and effectiveness. It is a reasonable assumption that the manner in which resources are used is more important than the actual quantity of available resources. The example of teacher aides has already been pointed out. Their actual function in the classroom can vary widely. The same is true for most pieces of equipment.

RECOMMENDATIONS

Because of the lack of conclusive results, no specific recommendations can be made concerning educational policy. However, a strong case can be made for questioning the historical acceptance of the precept that "more resources make a program better." Of course, resources are important in a program, and there are obviously minimum levels of resources. However, no relationship appears to be present between the general level of resources and the success of a program. Even when specific resources are examined, only the use of reading specialists in additional reading groups produces significant educational gains. The cost of special programs, however, is very high.

While the policy recommendations are limited, specific recommendations can be made regarding the course of future analyses of educational programs. These recommendations are the direct result of the problems encountered in the current study.

The first recommendation is that greater care must be taken in designing the measure of effectiveness. The greatest difficulty encountered in the current study centered around the measure of effectiveness and its lack of meaning when comparing classes with different pre-test scores. This

problem, represented by the heteroscedasticity in the measure of effectiveness, has reduced the confidence that can be placed in the study results. In fact, it may have masked any meaningful relationships.

Before any future studies are conducted, this problem should be investigated in an effort to achieve workable solutions. These solutions may include the use of correction factors on the residual gain scores, or the use of scaled scores rather than raw scores in the calculation of the residual gain scores. If these or other solutions are not feasible, consideration should be given to changing the analysis plan so the examination is restricted to a series of relatively homogeneous treatment groups.

Future analysis plans should be designed to ensure that series of definitive educational approaches form the basis for the analysis. The complexities and difficulties of achieving this goal are great. However, the lack of such a basis in this study reduced much of the analysis that could have been conducted.

If it is not practical to define meaningful educational approaches, consideration should be given to collecting resource and cost information that more accurately reflects actual resources provided to individuals in the classes. Our use of the special reading groups was an attempt in this direction. However, many situations occurred where we knew that special reading groups were being used, but they had not been identified as such.

It was demonstrated that policy decisions influenced the allocation resources to particular classes. The prime example of this situation was conventional classrooms with CR students only, which cost between 15 and 45 percent more than conventional classrooms with a mix of CR and NCR students. However, the CR students in both classes appeared to have similar abilities and both had approximately the same success. Therefore, certain policy decisions by the local education administrators must have influenced the allocation of funds.

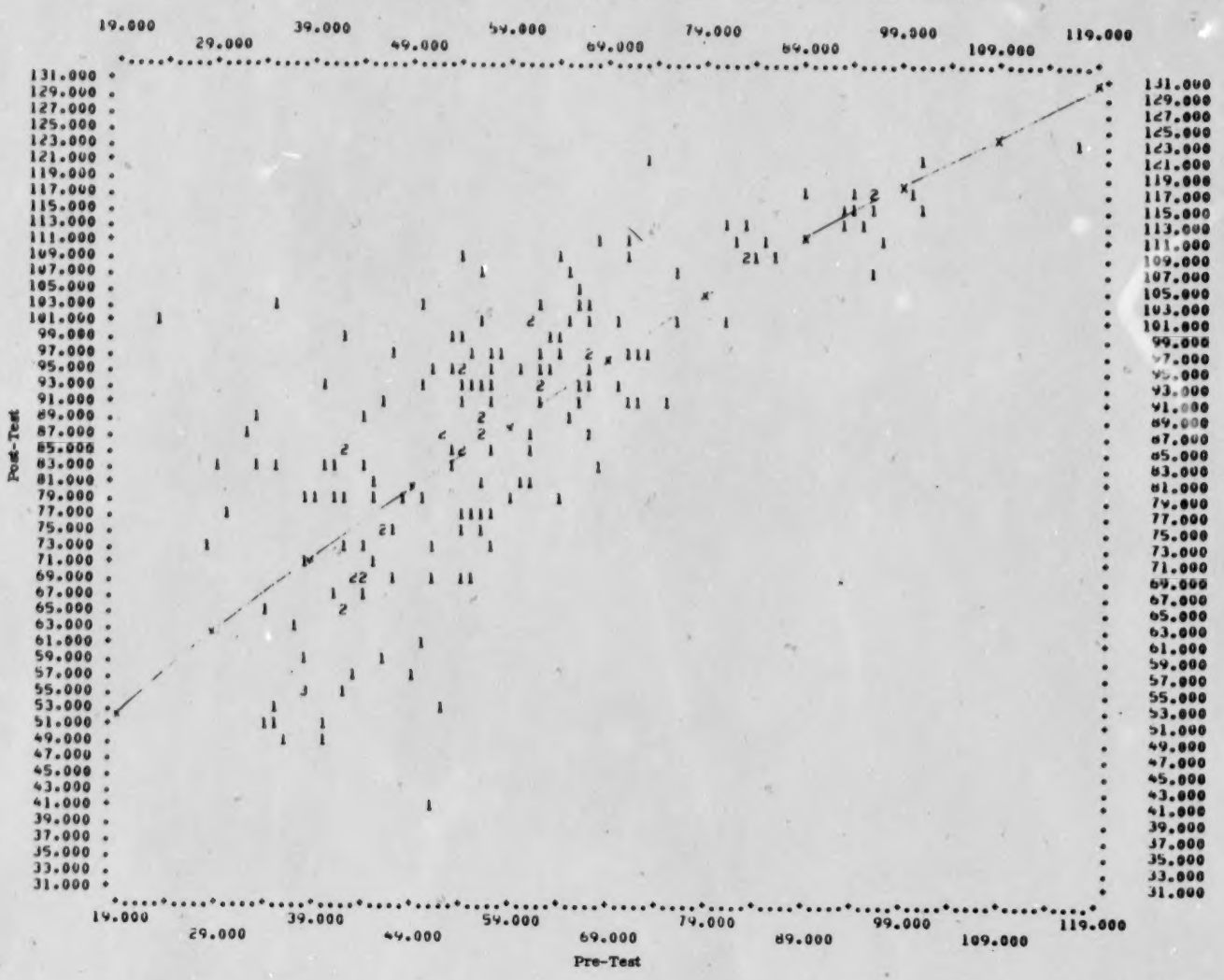
Future studies should attempt to examine these types of policy decisions in an effort to reduce their impact on the data. At the very least, future analyses should include an examination of the relationships between the socioeconomic levels of the students and the level of resource funding. This recommendation is closely associated with the previous recommendation regarding the identification of specific educational approaches. The policy decisions may, in fact, be embedded in particular educational approaches.

The last recommendation suggested as guidance for future studies is that program-effectiveness calculations should be based on the scores of students for whom both pre-test and post-test scores are available. While this will undoubtedly reduce the size of the available data base, it will significantly increase the confidence that can be placed in the resulting conclusions. The use of all available pre-test and post-test scores produces biases that are difficult to examine and may actually hide successful programs.

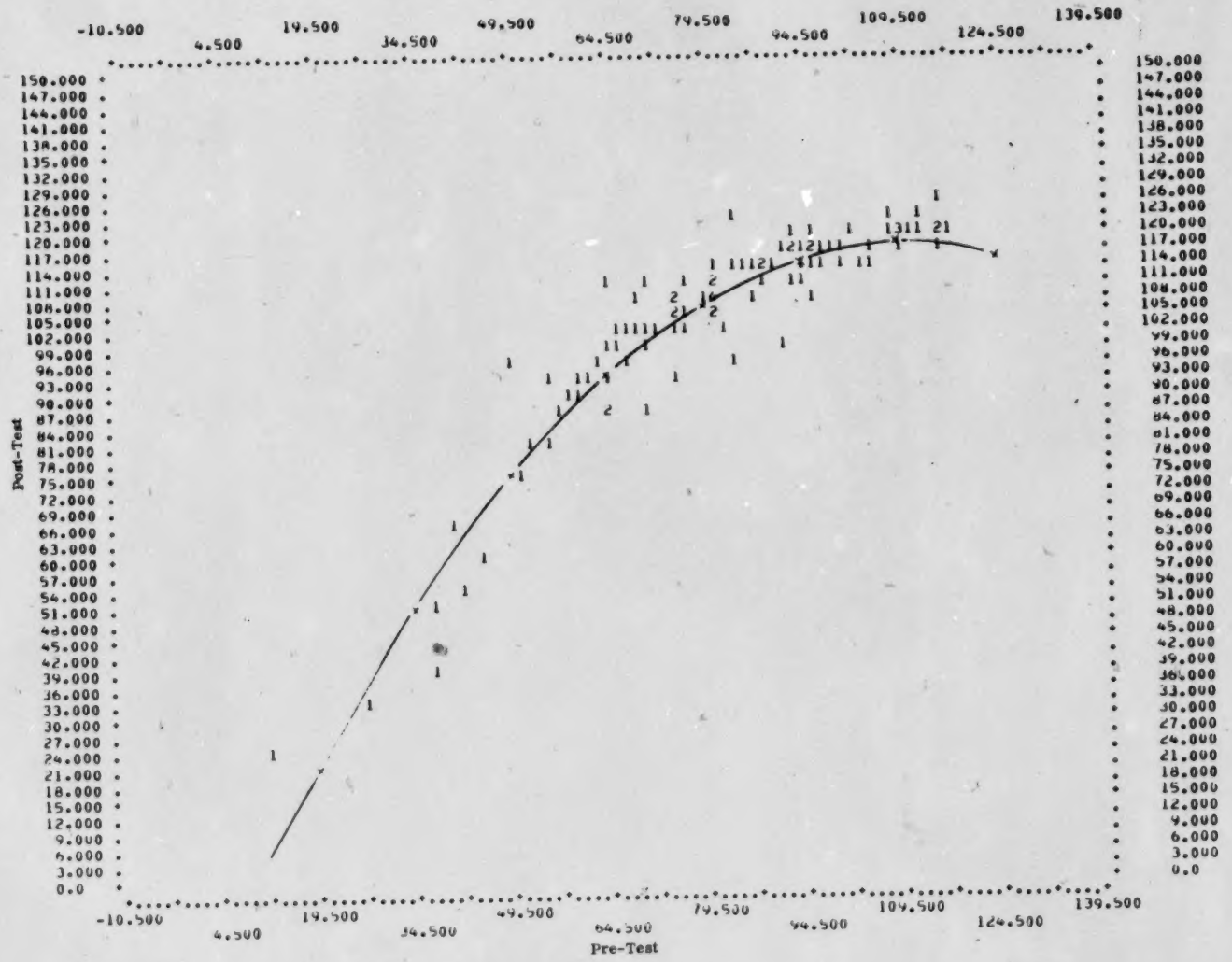
APPENDIX A

SCATTER DIAGRAMS OF PRE-TEST VERSUS POST-TEST SCORES

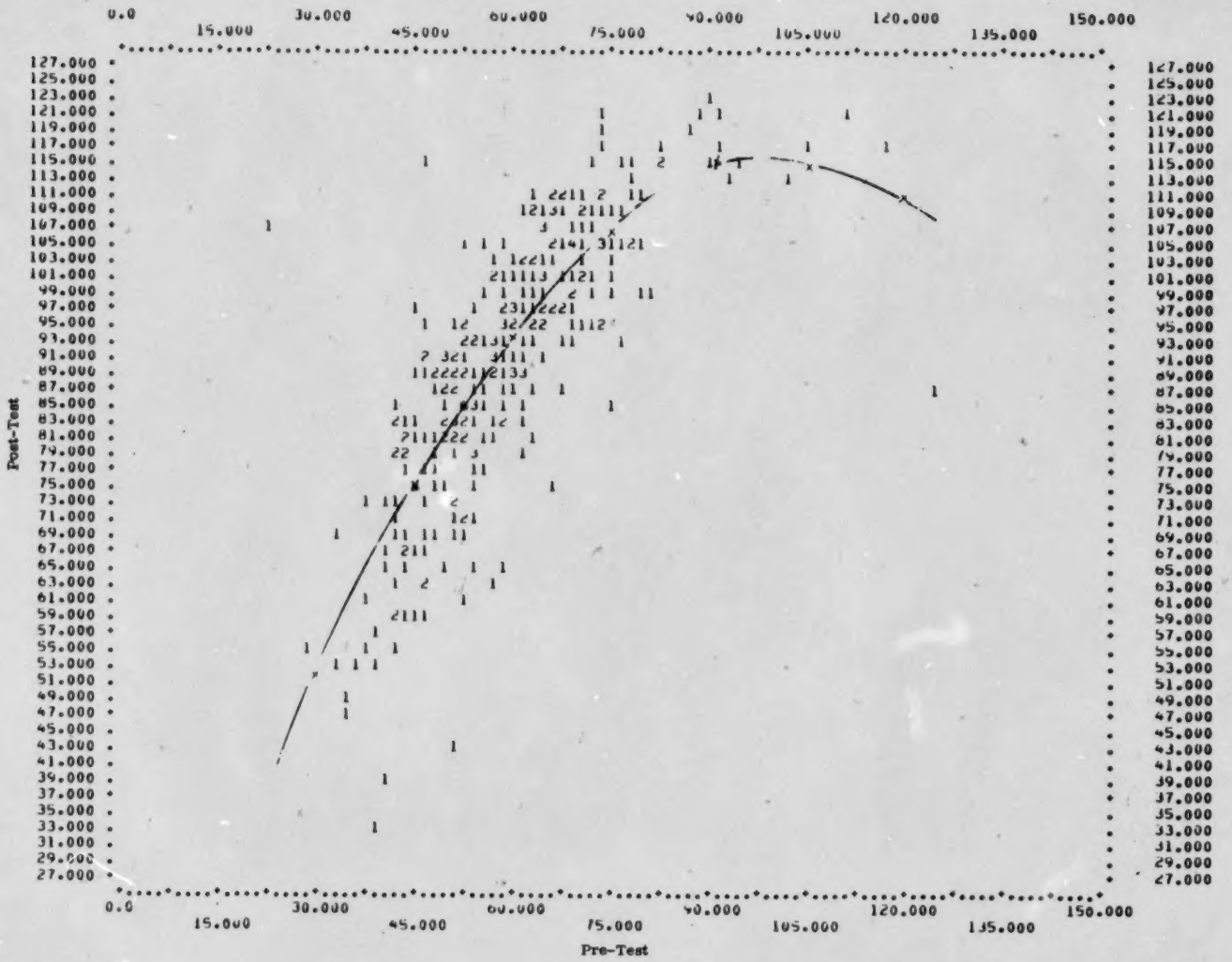
COMPENSATORY READING STUDENTS ONLY--GRADE 2



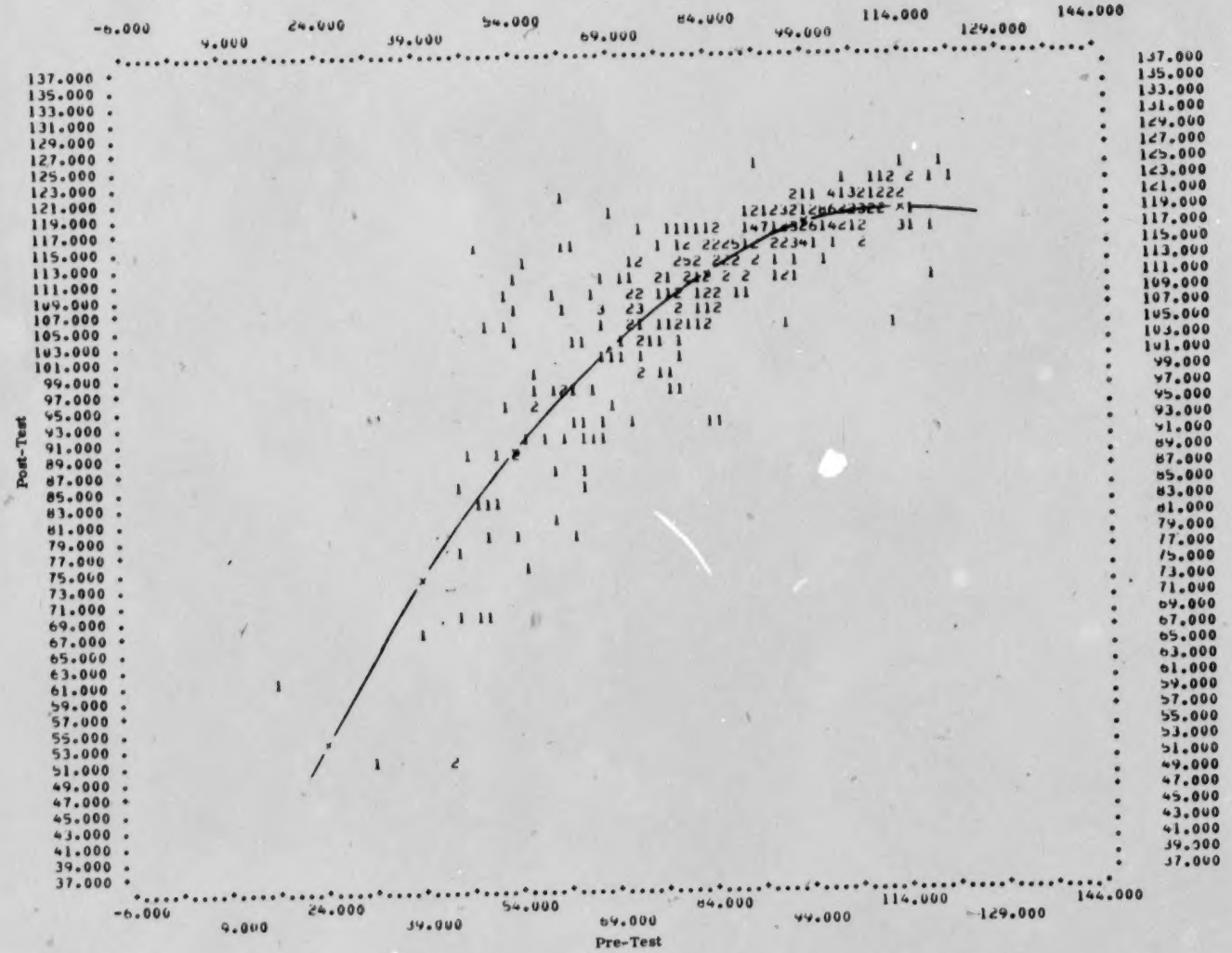
NONCOMPENSATORY READING STUDENTS ONLY--GRADE 2



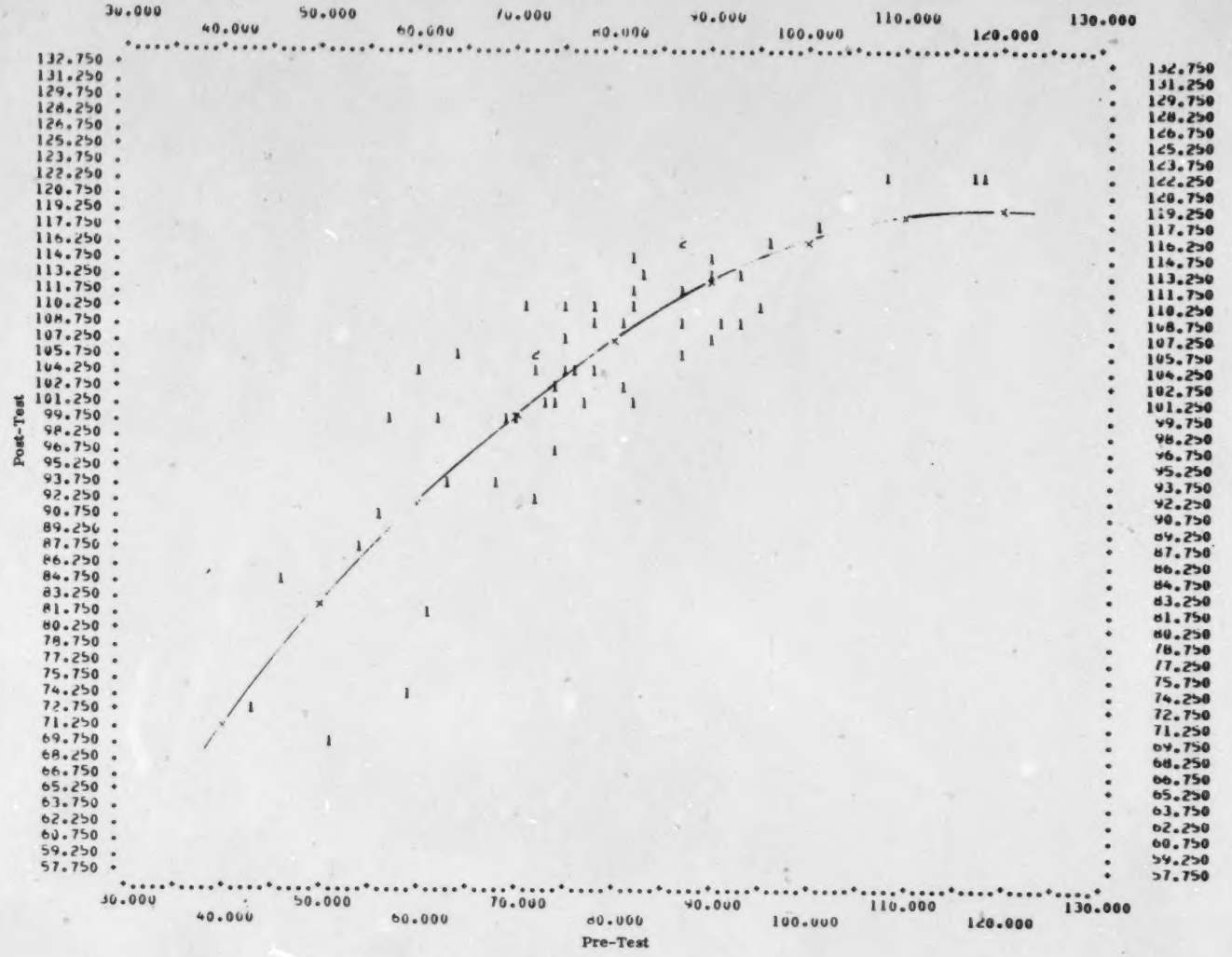
COMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 2



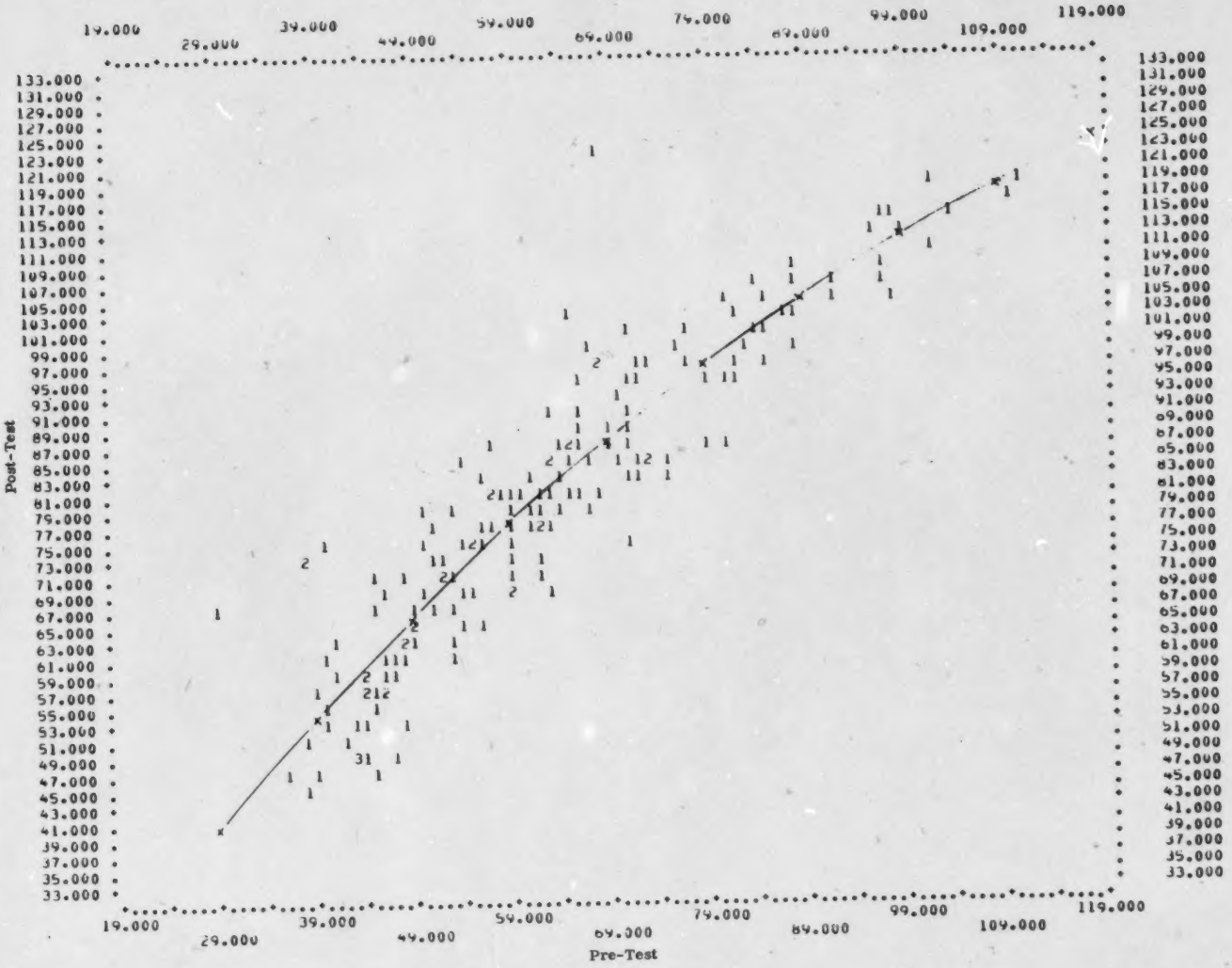
NONCOMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 2



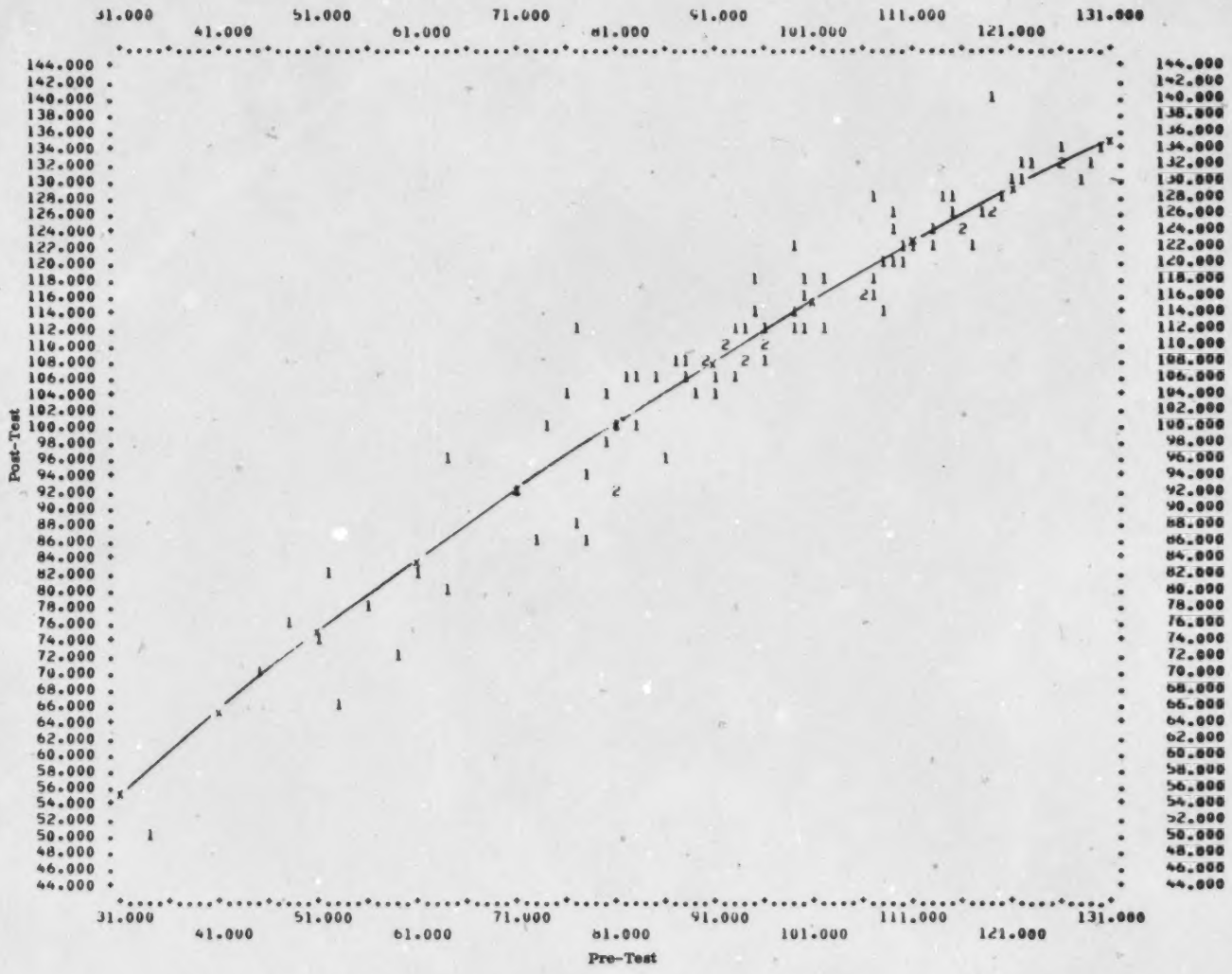
STUDENTS IN NONCOMPENSATORY READING SCHOOL--GRADE 2



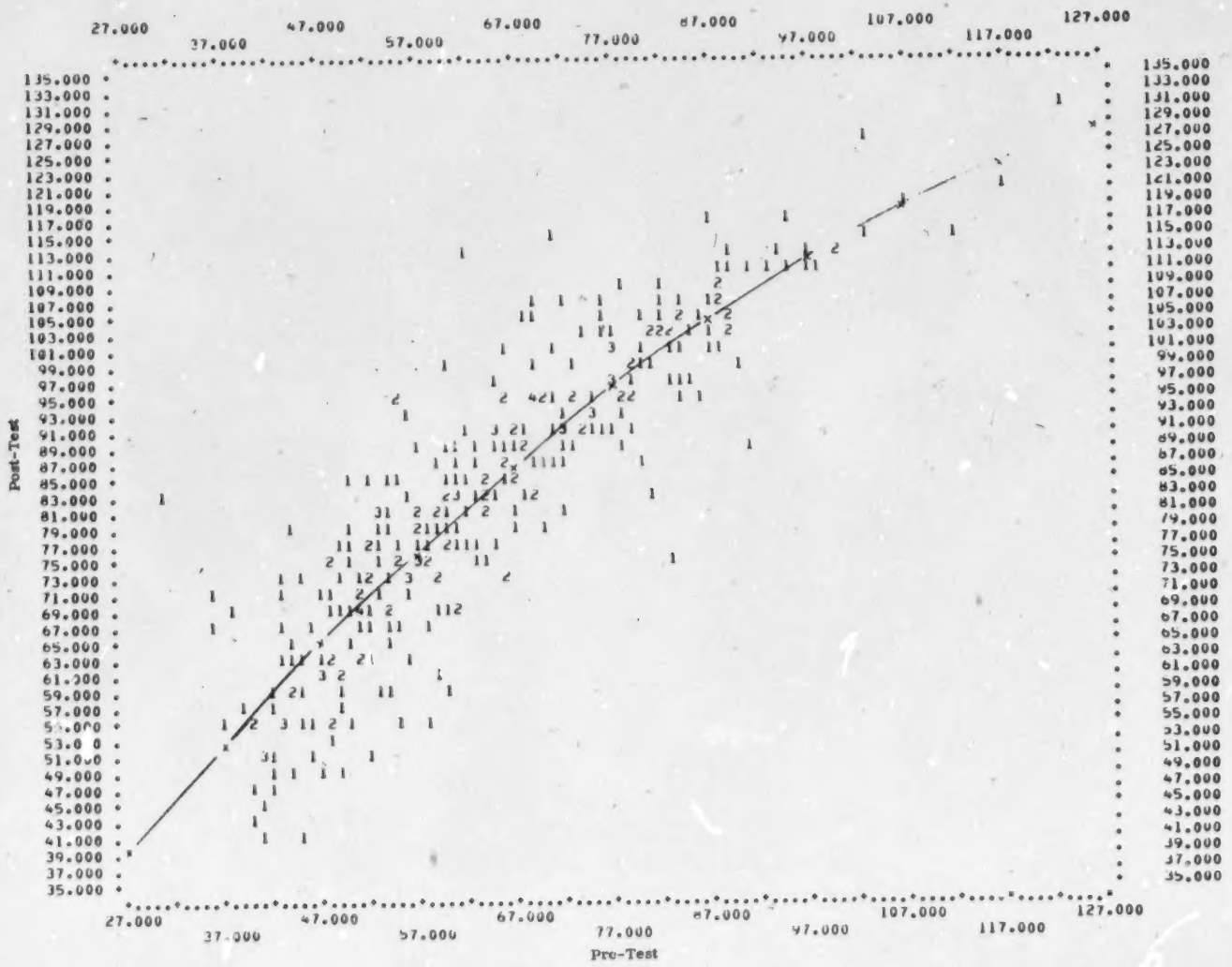
COMPENSATORY READING STUDENTS ONLY--GRADE 4



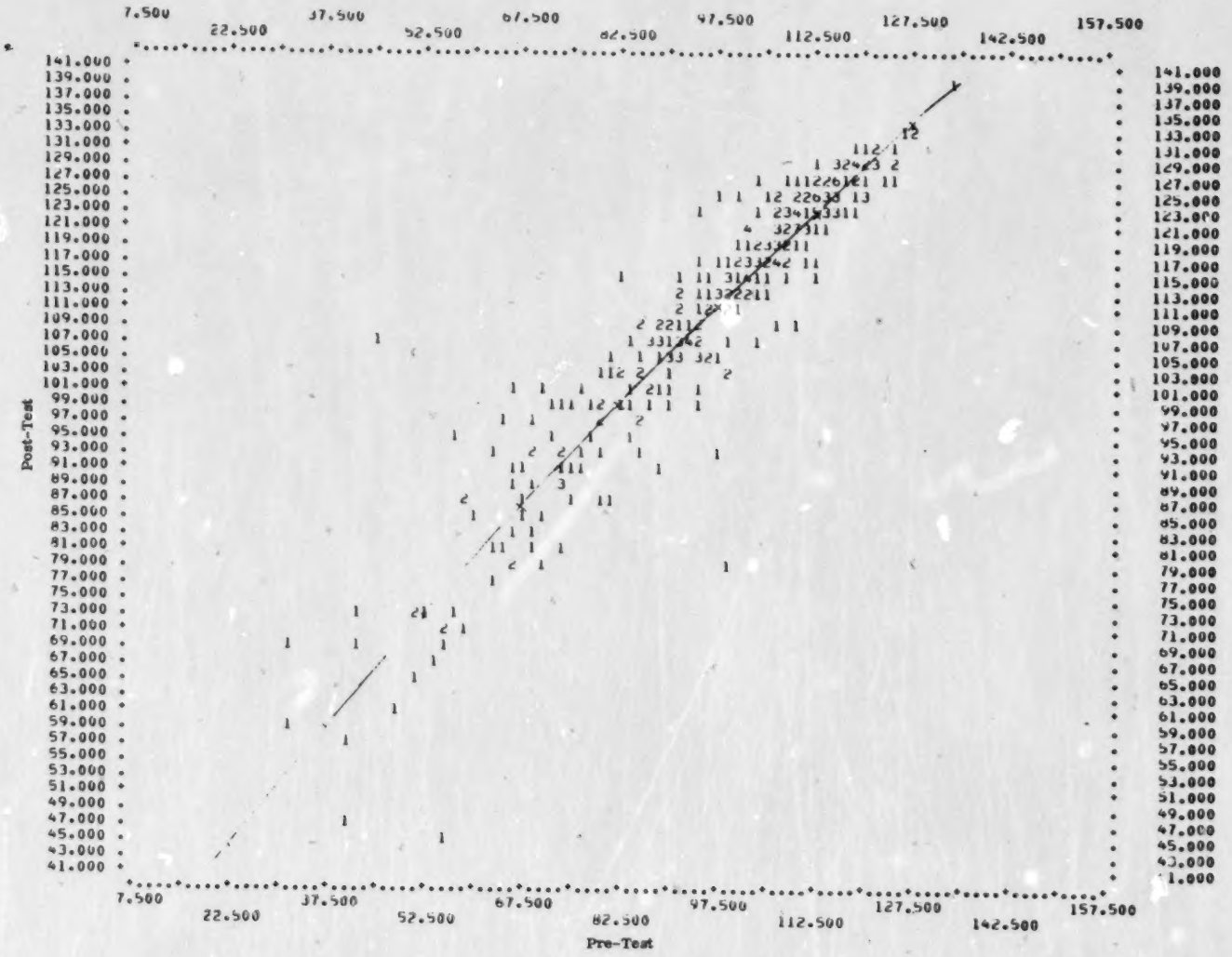
NONCOMPENSATORY READING STUDENTS ONLY--GRADE 4



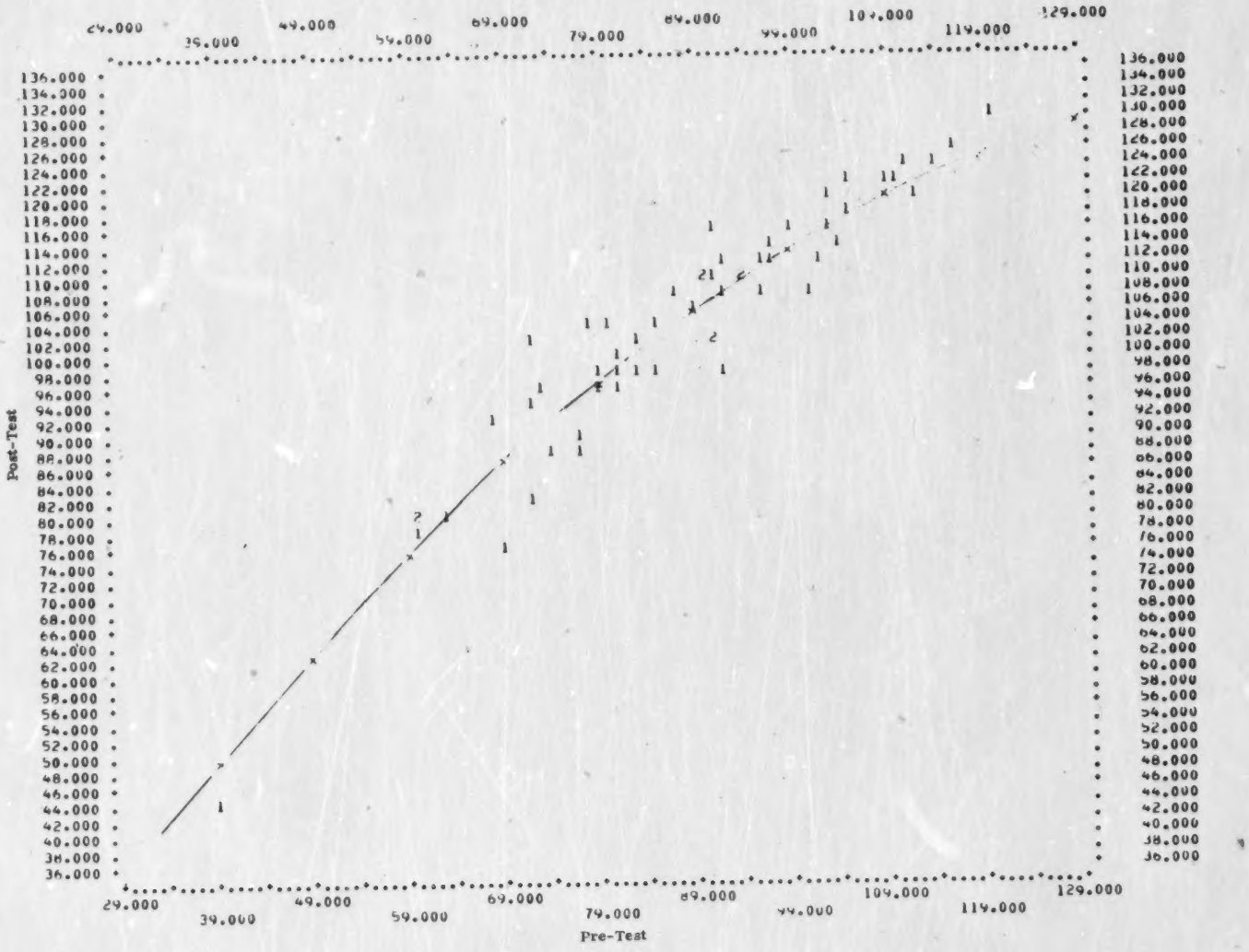
COMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 4



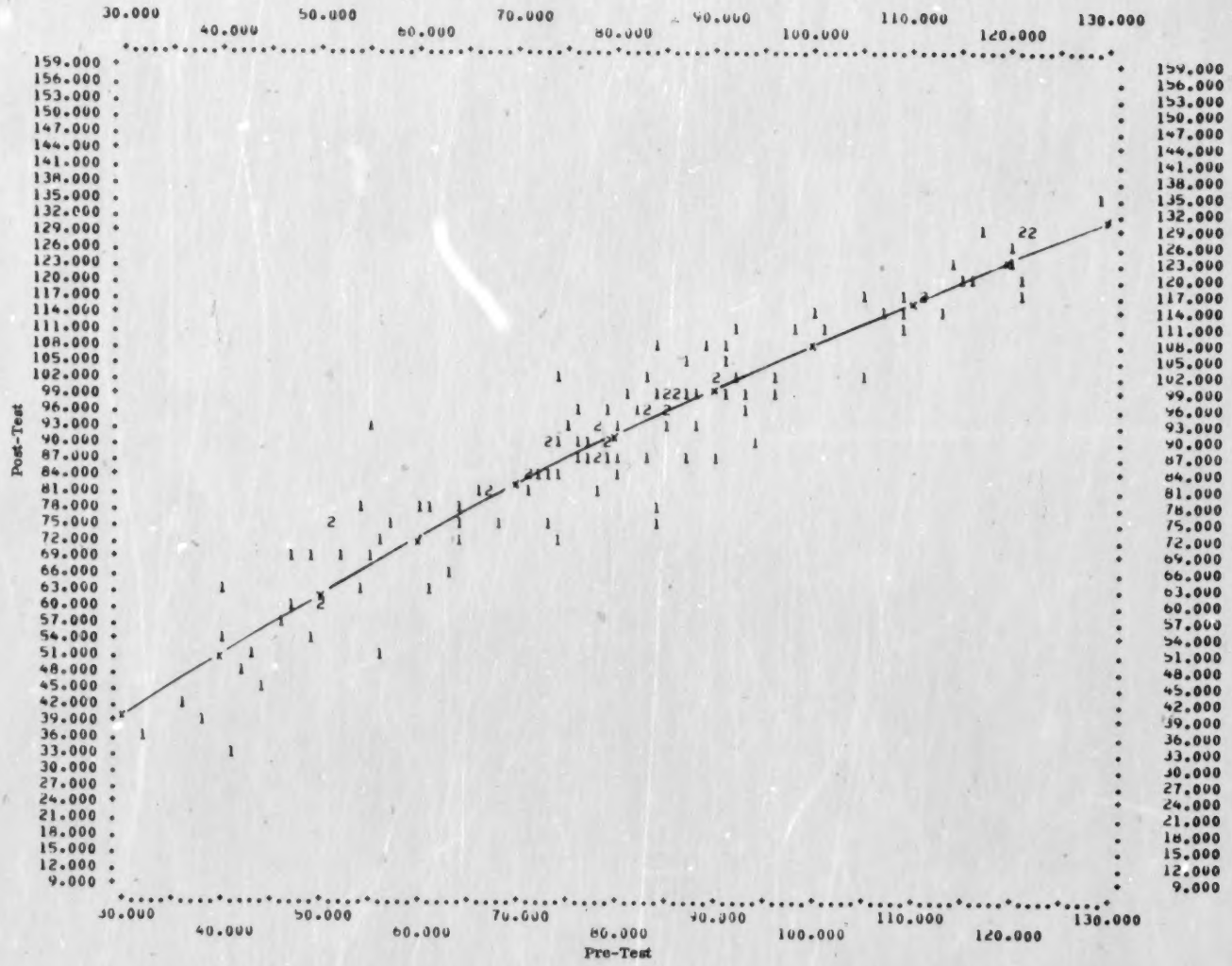
NONCOMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 4



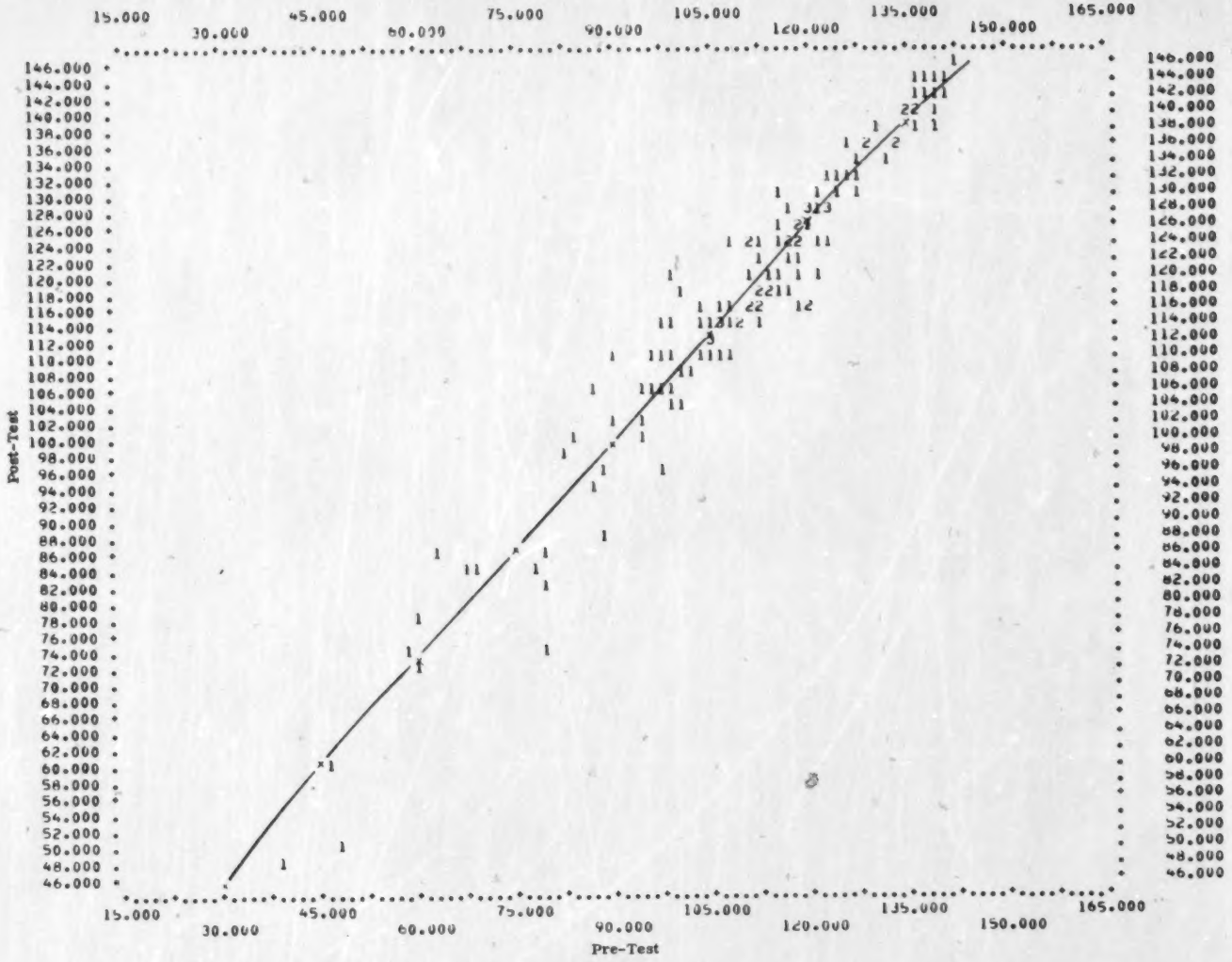
STUDENTS IN NONCOMPENSATORY READING SCHOOL--GRADE 4



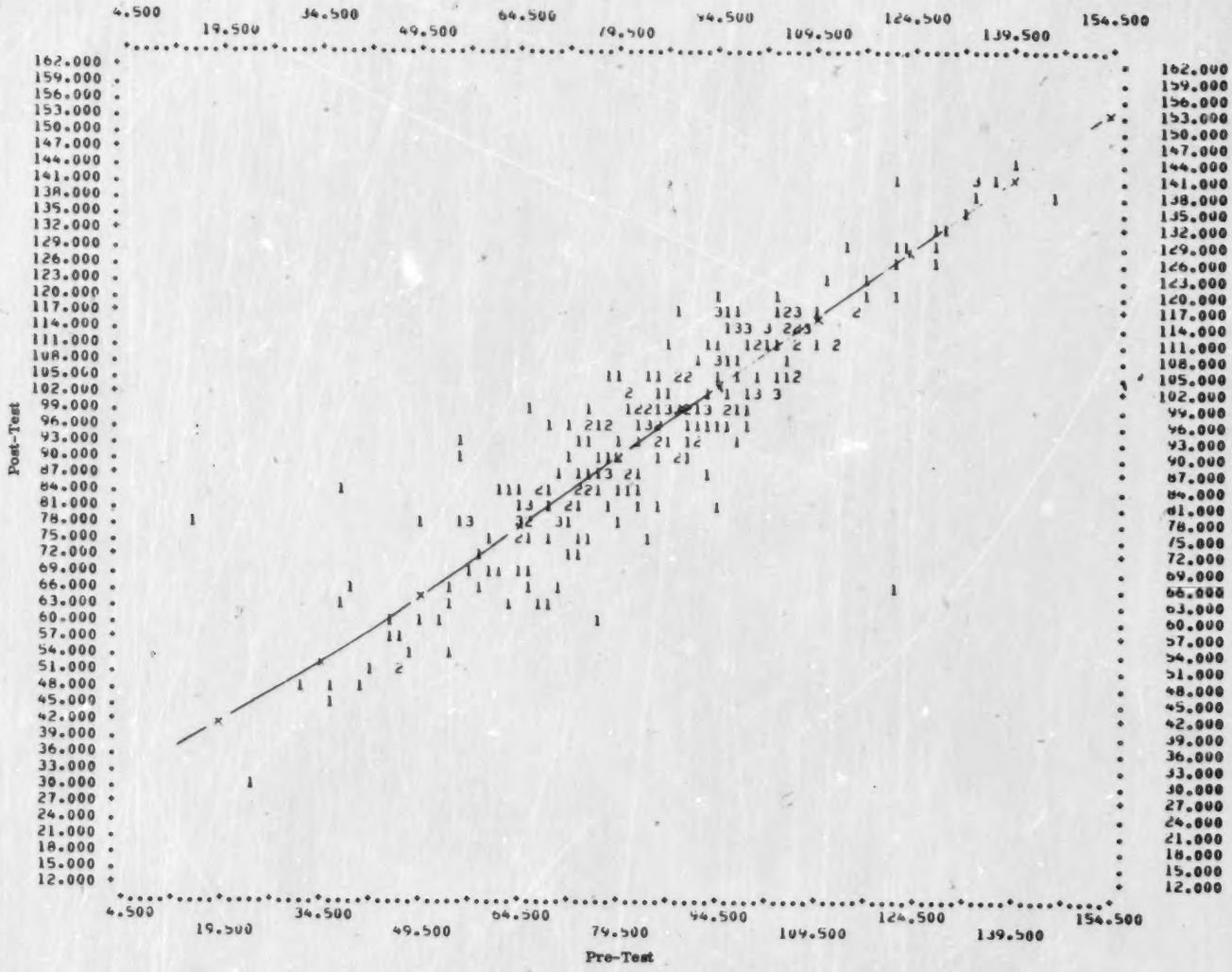
COMPENSATORY READING STUDENTS ONLY--GRADE 6



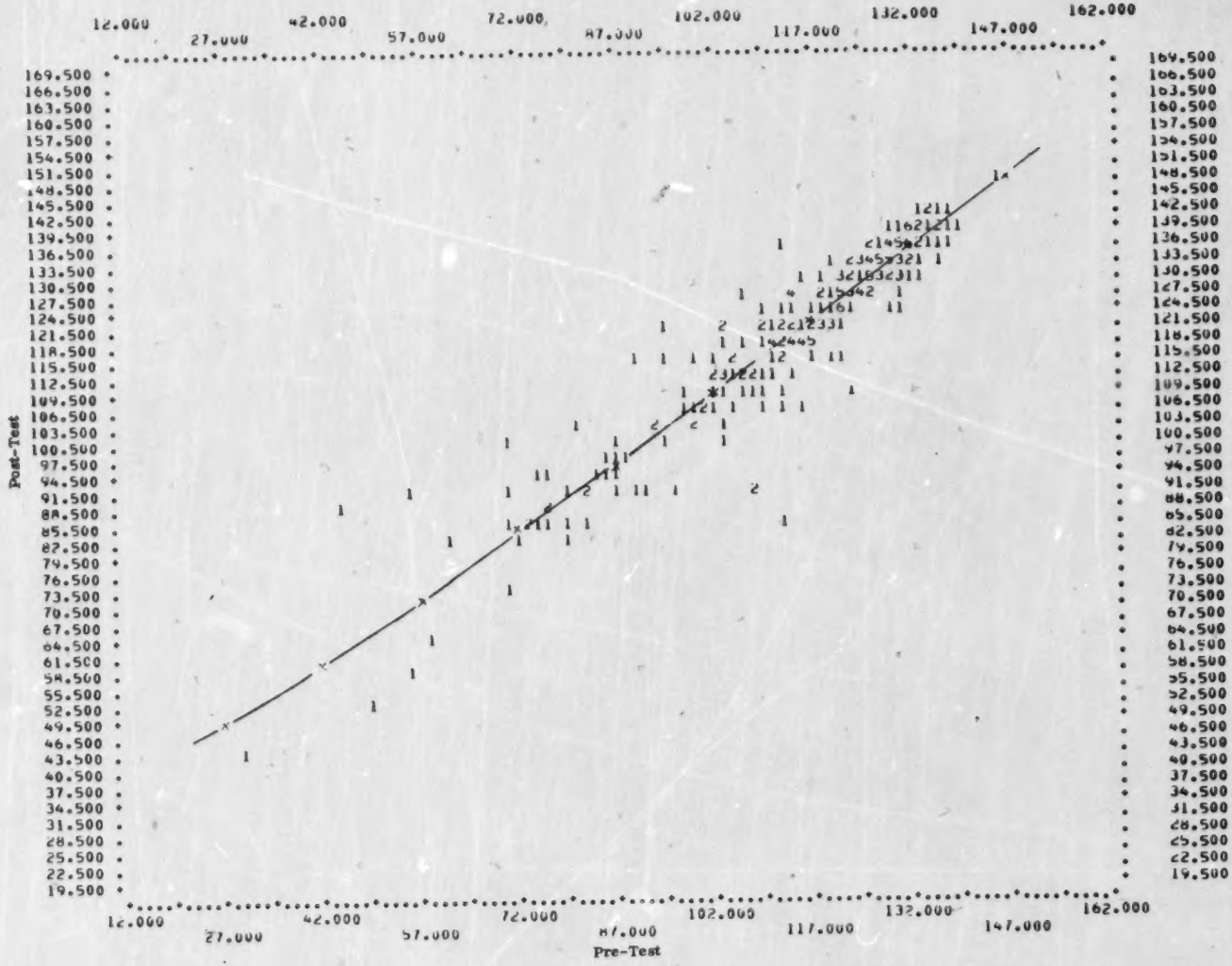
NONCOMPENSATORY READING STUDENTS ONLY--GRADE 6



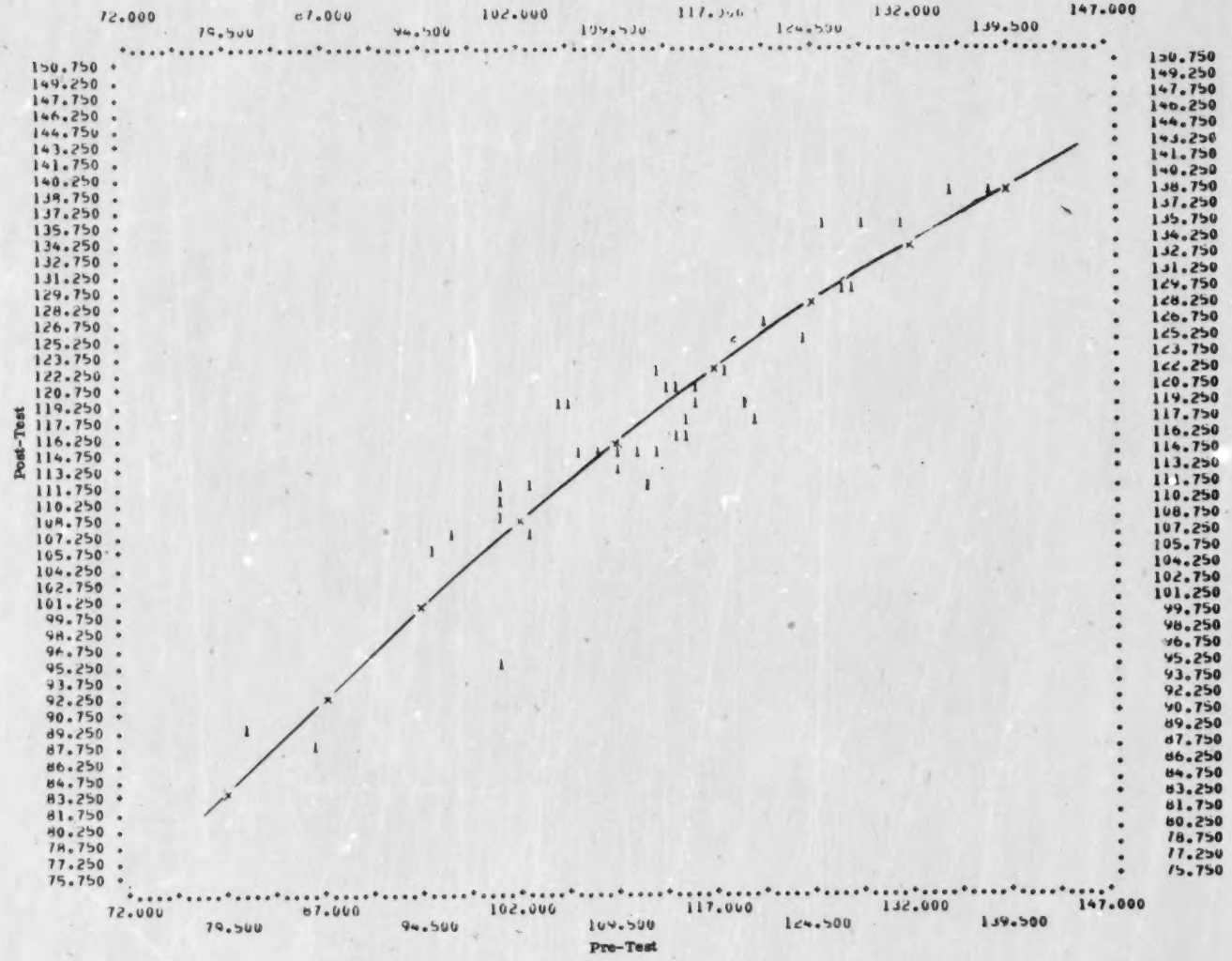
COMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 6



NONCOMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 6



STUDENTS IN NONCOMPENSATORY READING SCHOOL--GRADE 6



APPENDIX B

SIGNIFICANT RESULTS OF REGRESSION ANALYSIS
FOR THE QUADRATIC MEASURE OF EFFECTIVENESS

QUADRATIC MEASURE OF EFFECTIVENESS
COMPENSATORY READING STUDENTS ONLY--GRADE 2

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF CR STUDENTS	1.302	10.826

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	-4.56507	1.725	-2.647	1.000	0.0	
COST OF CLASSROOM INSTRUCTORS	0.05077	0.020	2.488	73.775	39.888	
COST OF TEACHER AIDES	-0.06999	0.027	-2.554	20.321	28.874	
COST OF SUPERVISION AND ADMINISTRATION	0.18924	0.030	6.342	21.575	27.030	
COST OF DISTRICT CENTERS	-0.17844	0.064	-2.787	6.328	12.075	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARED)	
R&R	0.3891
R&R-CORRECTED FOR DOF	0.3444
ABOUT THE MEAN	0.3599
ABOUT THE MEAN-CORRECTED FOR DOF	0.3359
MEAN OF DEPENDENT VARIABLE	1.302
STANDARD ERROR OF DEPENDENT VARIABLE	10.826
STANDARD ERROR OF ESTIMATE	8.796
F(5, 128) ABOUT ZERO	14.974
F(4, 128) ABOUT THE MEAN	17.589

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	5	5793.	1158.622
REGRESSION ABOUT MEAN	4	5568.	1391.896
DEVIATION	128	9904.	77.376
TOTAL	133	15697.	
TOTAL ABOUT MEAN	132	15472.	

QUADRATIC MEASURE OF EFFECTIVENESS
COMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 2

DEP. VARIABLE
EFFECTIVENESS OF CR STUDENTS

	MEAN	STD.DEV.
	0.865	7.680

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAN.COST*AVE.
CONSTANT	0.65468	0.801	0.817	1.000	0.0	
COST OF SLIDE/OVERHEAD/OPAQUE PROJECTORS	-3.63267	1.462	-2.485	0.235	0.344	
COST OF BASAL READERS	0.39404	0.197	2.005	2.702	2.562	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.0347
RAW-CORRECTED FOR DOF	0.0419
ABOUT THE MEAN	0.0426
ABOUT THE MEAN-CORRECTED FOR DOF	0.0340
MEAN OF DEPENDENT VARIABLE	0.865
STANDARD ERROR OF DEPENDENT VARIABLE	7.680
STANDARD ERROR OF ESTIMATE	7.548
F(3, 223) ABOUT ZERO	4.298
F(2, 223) ABOUT THE MEAN	4.984

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	3	735.	244.879
REGRESSION ABOUT MEAN	2	566.	282.818
DEVIATION	223	12706.	56.976
TOTAL	226	13440.	
TOTAL ABOUT MEAN	225	13271.	

132 146

QUADRATIC MEASURE OF EFFECTIVENESS
STUDENTS IN NONCOMPENSATORY READING SCHOOL--GRADE 2

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF NCR STUDENTS	0.181	5.649

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COSTWAVE.
CONSTANT	1.36562	0.811	1.684	1.000	0.0	
COST OF TEACHER AIDES	-0.09816	0.034	-2.910	11.457	21.402	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.1392
RAW-CORRECTED FOR DOF	0.1067
ABOUT THE MEAN	0.1383
ABOUT THE MEAN-CORRECTED FOR DOF	0.1220
MEAN OF DEPENDENT VARIABLE	0.181
STANDARD ERROR OF DEPENDENT VARIABLE	5.649
STANDARD ERROR OF ESTIMATE	5.293
F(2, 53) ABOUT ZERO	4.285
F(1, 53) ABOUT THE MEAN	8.506

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	2	240.	120.066
REGRESSION ABOUT MEAN	1	238.	238.331
DEVIATION	53	1485.	28.019
TOTAL	55	1725.	
TOTAL ABOUT MEAN	54	1723.	

QUADRATIC MEASURE OF EFFECTIVENESS
COMPENSATORY READING STUDENTS ONLY--GRADE 4

DEP. VARIABLE
EFFECTIVENESS OF CR STUDENTS

	MEAN	STD.DEV.
	-0.871	5.790

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	-1.32812	0.554	-2.398	1.000	0.0	
COST OF INDIVIDUAL READING KITS	1.57937	0.740	2.134	0.240	0.693	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.0572
RAW-CORRECTED FOR DOF	0.0419
ABOUT THE MEAN	0.0357
ABOUT THE MEAN-CORRECTED FOR DOF	0.0274
MEAN OF DEPENDENT VARIABLE	-0.871
STANDARD ERROR OF DEPENDENT VARIABLE	5.790
STANDARD ERROR OF ESTIMATE	5.709
F(2, 123) ABOUT ZERO	3.732
F(1, 123) ABOUT THE MEAN	4.555

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	2	243.	121.624
REGRESSION ABOUT MEAN	1	148.	148.449
DEVIATION	123	4009.	32.590
5.790 TOTAL	125	4252.	
TOTAL ABOUT MEAN	124	4157.	

QUADRATIC MEASURE OF EFFECTIVENESS
NONCOMPENSATORY READING STUDENTS ONLY--GRADE 4

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF NCR STUDENTS	-0.611	4.179

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	-2.111DC	0.870	-2.428	1.000	0.0	
COST OF FACILITIES	0.05721	0.029	1.997	26.228	15.380	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.0046
RAW-CORRECTED FOR DOF	0.0426
ABOUT THE MEAN	0.0443
ABOUT THE MEAN-CORRECTED FOR DOF	0.0332
MEAN OF DEPENDENT VARIABLE	-0.611
STANDARD ERROR OF DEPENDENT VARIABLE	4.179
STANDARD ERROR OF ESTIMATE	4.109

F(2, 86) ABOUT ZERO 2.268
F(1, 86) ABOUT THE MEAN 3.989

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	2	100.	50.121
REGRESSION ABOUT MEAN	1	67.	67.352
DEVIATION	86	1452.	16.885
TOTAL	88	1552.	
TOTAL ABOUT MEAN	87	1519.	

QUADRATIC MEASURE OF EFFECTIVENESS
COMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 4

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF CR STUDENTS	0.052	6.527

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	1.77894	0.931	1.910	1.000	0.0	
COST OF LIBRARY BOOKS	-0.18658	0.089	-2.097	9.254	4.896	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.0197
RAW-CORRECTED FOR DOF	0.0107
ABOUT THE MEAN	0.0196
ABOUT THE MEAN-CORRECTED FOR DOF	0.0151
MEAN OF DEPENDENT VARIABLE	0.052
STANDARD ERROR OF DEPENDENT VARIABLE	6.527
STANDARD ERROR OF ESTIMATE	6.477
F(2, 220) ABOUT ZERO	2.205
F(1, 220) ABOUT THE MEAN	4.345

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	2	185.	92.504
REGRESSION ABOUT MEAN	1	184.	184.349
DEVIATION	220	9230.	41.953
TOTAL	222	9415.	
TOTAL ABOUT MEAN	221	9414.	

QUADRATIC MEASURE OF EFFECTIVENESS
NONCOMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 4

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF NCR STUDENTS	0.308	4.706

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	-0.70415	0.485	-1.453	1.000	0.0	
COST OF NON-TEACHING NON-ACADEMIC STAFF	0.61322	0.203	3.013	1.341	1.341	
COST OF DISTRICT CENTERS	-0.11651	0.043	-2.706	2.872	6.298	
COST OF BASAL READERS	0.23742	0.115	2.060	2.209	2.362	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.0615
RAW-CORRECTED FOR DOF	0.0484
ABOUT THE MEAN	0.0574
ABOUT THE MEAN-CORRECTED FOR DOF	0.0476
MEAN OF DEPENDENT VARIABLE	0.308
STANDARD ERROR OF DEPENDENT VARIABLE	4.706
STANDARD ERROR OF ESTIMATE	4.593
F(4, 287) ABOUT ZERO	4.700
F(3, 287) ABOUT THE MEAN	5.830

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	4	397.	99.131
REGRESSION ABOUT MEAN	3	369.	122.956
DEVIATION	287	6053.	21.092
TOTAL	291	6450.	
TOTAL ABOUT MEAN	290	6422.	

QUADRATIC MEASURE OF EFFECTIVENESS
STUDENTS IN NONCOMPENSATORY READING SCHOOL--GRADE 4

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF NCR STUDENTS	0.244	4.429

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	-1.36511	0.710	-1.921	1.000	0.0	
COST OF TELEVISIONS/VIDEO TAPE RECORDERS	1.41813	0.548	2.586	0.461	0.998	
COST OF DISTRICT CENTERS	0.32663	0.127	2.569	2.426	4.304	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.1954
RAW-CORRECTED FOR DOF	0.1499
ABOUT THE MEAN	0.1929
ABOUT THE MEAN-CORRECTED FOR DOF	0.1625
MEAN OF DEPENDENT VARIABLE	0.244
STANDARD ERROR OF DEPENDENT VARIABLE	4.429
STANDARD ERROR OF ESTIMATE	4.053
F(3, 53) ABOUT ZERO	4.291
F(2, 53) ABOUT THE MEAN	6.334

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	3	211.	70.490
REGRESSION ABOUT MEAN	2	208.	104.069
DEVIATION	53	871.	16.429
TOTAL	56	1082.	
TOTAL ABOUT MEAN	55	1079.	

QUADRATIC MEASURE OF EFFECTIVENESS
COMPENSATORY READING STUDENTS ONLY--GRADE 6

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF CR STUDENTS	0.657	5.340

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	2.02939	0.662	3.067	1.000	0.0	
COST OF TEACHING MACHINES	-1.98434	0.895	-2.218	0.237	0.580	
COST OF DISTRICT CENTERS	-0.19388	0.074	-2.613	4.659	6.996	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)	
RAW	0.1187
RAW*CORRECTED FOR DOF	0.0909
ABOUT THE MEAN	0.1053
ABOUT THE MEAN*CORRECTED FOR DOF	0.0864
MEAN OF DEPENDENT VARIABLE	0.657
STANDARD ERROR OF DEPENDENT VARIABLE	5.340
STANDARD ERROR OF ESTIMATE	5.105

F(3, 95) ABOUT ZERO 4.200
F(2, 95) ABOUT THE MEAN 5.588

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	3	333.	111.154
REGRESSION ABOUT MEAN	2	291.	145.598
DEVIATION	95	2475.	26.056
TOTAL	98	2809.	
TOTAL ABOUT MEAN	97	2767.	

QUADRATIC MEASURE OF EFFECTIVENESS
COMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 6

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF CR STUDENTS	0.283	0.543

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST*AVE.
CONSTANT	-5.21243	2.179	-2.392	1.000	0.0	
COST OF OTHER SUPPLIES AND MATERIALS	0.36456	2.464	2.583	0.863	0.192	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)

RAW	0.0360
RAW-CORRECTED FOR DOF	0.0262
ABOUT THE MEAN	0.0348
ABOUT THE MEAN-CORRECTED FOR DOF	0.0296
MEAN OF DEPENDENT VARIABLE	0.283
STANDARD ERROR OF DEPENDENT VARIABLE	0.543
STANDARD ERROR OF ESTIMATE	0.445

F(2, 185) ABOUT ZERO	3.510
F(1, 185) ABOUT THE MEAN	6.671

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	2	292.	146.052
REGRESSION ABOUT MEAN	1	277.	277.105
DEVIATION	185	7685.	41.538
0.543 TOTAL	187	7977.	
TOTAL ABOUT MEAN	186	7962.	

QUADRATIC MEASURE OF EFFECTIVENESS
NONCOMPENSATORY READING STUDENTS IN MIXED CLASSES--GRADE 6

DEP. VARIABLE	MEAN	STD.DEV.
EFFECTIVENESS OF NCR STUDENTS	0.337	4.201

VARIABLE	COEFFICIENT	STD.ERROR	T-VALUE	MEAN	STD.DEV.	MAR.COST#AVE.
CONSTANT	0.06136	0.309	0.198	1.000	0.0	
COST OF OTHER SPECIAL TEACHERS	0.04490	0.022	2.068	6.137	13.044	

DIAGNOSTIC STATISTICS FOR THE REGRESSION

COEFFICIENTS OF DETERMINATION (R-SQUARES)

RAW	0.0257
RAW.CORRECTED FOR DOF	0.0169
ABOUT THE MEAN	0.0194
ABOUT THE MEAN.CORRECTED FOR DOF	0.0150
MEAN OF DEPENDENT VARIABLE	0.337
STANDARD ERROR OF DEPENDENT VARIABLE	4.201
STANDARD ERROR OF ESTIMATE	4.170

F(2, 220) ABOUT ZERO 2.904
F(1, 220) ABOUT THE MEAN 4.359

ANALYSIS OF VARIANCE FOR THE REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES
REGRESSION	2	101.	50.494
REGRESSION ABOUT MEAN	1	76.	75.793
DEVIATION	220	3825.	17.388
TOTAL	222	3926.	
TOTAL ABOUT MEAN	221	3901.	

APPENDIX C

RESULTS OF THE ANALYSIS OF VARIANCE
OF THE COSTS OF EFFECTIVE PROGRAMS

COST OF SLIDE PROJECTORS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed					Mixed					Mixed		
	CR	NCR	CR	NCR	NCR SCH	CR	NCR	CR	NCR	NCR SCH	CR	NCR	CR	NCR	NCR SCH
Average Cost for Effective Programs	.18	.46	.35	.19	.57	1.00	.33	.14	.23	.22	.12	.28	.12	.30	.17
Average Cost for Average Programs	.59	.35	.26	.28	.46	.35	.25	.30	.22	.39	.32	.34	.32	.52	.36
Average Cost for Ineffective Programs	.18	.02	.33	.67	.08	.49	.29	1.11	1.0	.25	.42	.21	.42	.23	.28
F Test	1.74	4.22[a]	1.05	5.56[a]	2.39	3.95[a]	.62	2.07	.19	.70	2.70	4.60[a]	1.09	1.47[a]	8.82[a]
R^2	.03	.04	.01	.04	.03	.03	.02	.02	.00	.03	.05	.09	.01	.13	.33

a. Significant differences exist at the 0.05 level.

COST OF OTHER STAFF IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR SCH			Mixed		NCR SCH			Mixed		NCR SCH
	CR	NCR	CR	NCR		CR	NCR	CR	NCR		CR	NCR	CR	NCR	
Average Cost for Effective Programs	23.32	26.50	23.95	41.35	21.50	28.40	29.12	15.00	24.71	19.78	9.78	15.79	9.78	46.55	10.67
Average Cost for Average Programs	28.26	16.61	19.23	19.63	12.90	27.36	19.80	22.22	24.08	18.18	22.74	13.10	22.74	21.89	15.15
Average Cost for Ineffective Programs	5.93	18.25	31.29	26.41	14.64	48.12	11.83	19.81	30.08	10.14	17.92	11.83	17.92	28.82	10.71
F Test	25.99[a]	.20	.65	5.01[a]	.60	.52	1.08	.40	.24	.24	3.85[a]	.51	17.4	3.34[a]	.75
R^2	.21	0	0	.03	.02	.01	.02	0	0	.01	.06	.01	.09	.03	.04

a. **Significant differences exist at the 0.05 level.**

COST OF READING SPECIALISTS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR			Mixed		NCR			Mixed		NCR
	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH
Average Cost for Effective Programs	150.81	50.25	98.18	136.31	37.00	412.00	49.50	70.34	8.89	NA	34.00	58.56	34.00	49.40	16.00
Average Cost for Average Programs	111.17	38.24	52.16	42.84	28.82	72.51	55.37	66.85	161.62	NA	57.02	29.94	57.02	62.39	19.11
Average Cost for Ineffective Programs	100.93	37.75	158.60	90.80	29.43	471.57	46.00	74.27	284.50	NA	29.20	32.00	29.20	137.00	16.50
F Test	2.73	.07	6.26[a]	4.95[a]	.10	7.03[a]	.05	.04	6.20[a]	NA	.18	.10	.45	4.33[a]	1.60
R^2	.04	0	.08	.01	.01	.11	0	0	.07	NA	0	0	.01	.07	.23

a. **Significant differences exist at the 0.05 level.**

COST OF TEACHER AIDES IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed					Mixed					Mixed		
	CR	NCR	CR	NCR	NCR SCH	CR	NCR	CR	NCR	NCR SCH	CR	NCR	CR	NCR	NCR SCH
Average Cost for Effective Programs	47.35	4.66	29.50	16.00	29.44	51.00	3.60	12.25	22.38	17.00	4.20	6.69	4.20	9.33	2.50
Average Cost for Average Programs	32.95	14.13	12.84	10.36	15.47	32.23	8.75	15.09	14.49	11.67	17.74	8.35	17.74	15.64	19.67
Average Cost for Ineffective Programs	22.95	19.00	43.25	12.80	12.00	37.67	4.00	13.40	28.56	16.33	7.50	6.50	7.50	15.80	18.50
F Test	1.53	.50	6.99[a]	.26	1.36	.15	1.46	.11	5.26[a]	1.03	.26	.16	.65	.06	.95
R^2	.03	.02	.07	0	.01	0	.07	0	.06	.01	0	.01	.08	0	.10

a. Significant differences exist at the 0.05 level.

COST OF CLASSROOM INSTRUCTOR IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR			Mixed		NCR			Mixed		NCR
	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH
Average Cost for Effective Programs	55.24	75.00	61.99	66.52	75.71	54.40	41.00	56.94	67.22	38.60	66.11	48.00	66.11	57.17	66.54
Average Cost for Average Programs	73.26	76.55	66.04	44.91	82.60	7.68	10.19	61.98	61.50	62.72	26.55	72.10	26.55	52.89	68.52
Average Cost for Ineffective Programs	94.11	117.33	66.25	56.29	64.17	82.00	63.00	46.91	54.57	47.38	52.69	47.56	52.69	56.50	67.93
F Test	13.04[a]	.85	.11	2.00	1.22	21.50[a]	.08	1.05	.31	11.58[a]	1.74	1.72	1.03	.32	.10
R^2	.13	.02	0	.10	.04	.05	0	.01	0	.30	.03	.03	.13	0	.01

a. **Significant differences exist at the 0.05 level.**

COST OF LIBRARY BOOKS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed					Mixed					Mixed		
	CR	NCR	CR	NCR	NCR SCH	CR	NCR	CR	NCR	NCR SCH	CR	NCR	CR	NCR	NCR SCH
Average Cost for Effective Programs	10.29	9.77	9.67	12.26	14.40	12.62	12.64	10.19	10.19	11.62	12.86	4.65	8.67	10.54	6.83
Average Cost for Average Programs	9.67	11.05	9.29	9.43	8.79	12.06	8.00	10.86	10.86	10.16	12.31	7.97	9.65	10.27	8.80
Average Cost for Ineffective Programs	7.43	7.39	15.47	12.82	10.00	32.44	5.42	11.38	11.38	6.75	13.13	7.00	6.92	15.97	8.38
F Test	.26	.14	54.44[a]	5.75	4.20[a]	3.19[a]	1.51	1.50	.85	1.55	.17	6.07[a]	1.15	7.84[a]	.22
R^2	0	0	.20	.04	.16	.04	.03	0	.01	.06	.00	.09	.01	.06	.01

a. Significant differences exist at the 0.05 level.

COST OF INDIVIDUAL READING KITS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR			Mixed		NCR			Mixed		NCR
	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH
Average Cost for Effective Programs	.75	.73	.91	.79	1.00	.65	.29	.52	.25	.43	.25	.10	.25	.38	.09
Average Cost for Average Programs	.77	.41	.71	.68	.75	.61	.64	.53	.49	.43	.59	.80	.59	.66	.35
Average Cost for Ineffective Programs	1.23	1.00	.77	1.23	.47	.80	.20	.50	2.67	1.00	.85	.77	.86	.40	.53
F Test	.73	3.84[a]	.35	3.28[a]	2.64	2.70	1.55	.13	10.40[a]	.70	.05	2.67	.33	.30	.86
R^2	.02	.14	0	.05	.03	.07	.08	0	.12	.08	0	.07	0	0	.11

a. Significant differences exist at the 0.05 level.

COST OF BASAL READERS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR SCH			Mixed		NCR SCH			Mixed		NCR SCH
	CR	NCR	CR	NCR		CR	NCR	CR	NCR		CR	NCR	CR	NCR	
Average Cost for Effective Programs	2.41	3.54	3.25	3.25	3.00	2.04	2.86	2.64	1.38	1.60	2.00	1.84	2.00	5.78	1.67
Average Cost for Average Programs	3.85	3.09	3.15	.49	1.24	3.81	2.58	2.64	3.04	1.92	1.98	2.65	1.98	4.52	3.58
Average Cost for Ineffective Programs	1.95	3.40	3.73	2.73	2.27	3.63	3.82	2.27	2.64	3.43	2.15	1.53	2.15	2.92	3.37
F Test	2.46	.09	.17	4.05[a]	.96	7.03[a]	.90	.10	1.87	1.08	1.77	1.53	.02	3.75[a]	.40
R^2	.03	.02	0	.03	.04	.09	.13	0	.01	.04	.03	.03	0	.06	.02

a. Significant differences exist at the 0.05 level.

COST OF MOVIE PROJECTORS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR			Mixed		NCR			Mixed		NCR
	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH
Average Cost for Effective Programs	1.43	.50	1.32	1.00	1.07	1.60	.69	.50	.66	.88	.87	.74	.88	1.00	1.17
Average Cost for Average Programs	1.23	.85	1.15	1.03	3.38	1.73	.46	.98	.79	1.03	.83	.66	.83	1.09	.81
Average Cost for Ineffective Programs	1.00	1.00	1.31	.88	2.00	2.14	.36	.76	.79	.88	1.00	.61	1.00	1.00	1.57
F Test	.36	.31	.62	.11	7.94[a]	.26	1.79	.98	.09	.09	.89	.08	.33	1.61	2.31
R^2	0	0	0	0	.25	0	.04	.01	0	0	.02	0	0	.01	.11

a. Significant differences exist at the 0.05 level.

COST OF TAPE RECORDERS IN DOLLARS PER STUDENT

	Grade 2					Grade 4					Grade 6				
			Mixed		NCR			Mixed		NCR			Mixed		NCR
	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH	CR	NCR	CR	NCR	SCH
Average Cost for Effective Programs	1.86	1.67	0	2.26	2.79	1.00	1.00	1.50	1.25	2.25	.88	1.00	.56	1.09	N/A
Average Cost for Average Programs	2.44	1.76	6.38	1.75	2.10	2.80	1.16	1.37	1.45	1.68	1.41	1.18	1.41	1.27	N/A
Average Cost for Ineffective Programs	3.43	1.67	2.83	3.12	2.25	2.33	3.64	2.00	2.29	1.50	1.83	.72	1.83	1.57	N/A
F Test	.95	.03	4.40[a]	6.09[a]	1.10	2.00	.50	1.19	3.12[a]	.44	.05	.84	1.52	.17	N/A
R^2	.01	0	.03	.04	.04	.02	.04	.01	.01	.02	0	.01	.01	0	N/A

a. Significant differences exist at the 0.05 level.